Universidad Zaragoza
1542

Trabajo Fin de Máster

Movilidad IP en redes heterogéneas: optimización de flujos de tráfico con QoS

Autor
Andrés Chesa Badia

Director
Jose María Saldaña Medina

Ponente
José Ruiz Mas

Universidad de Zaragoza
Escuela de Ingeniería y Arquitectura

2013

Resumen

El incremento en la demanda de la tasa de bit para satisfacer la calidad de la experiencia en los servicios inalámbricos que suministra la infraestructura de red al usuario, tiene como consecuencia una evolución y mejora en las redes heterogéneas de nueva generación (HetNets). El nuevo modelo está caracterizado por la coexistencia de diferentes estándares Wireless, UMTS junto al 3GPP LTE y el 802.11. Cada estándar tiene diferentes características en términos de cobertura, gestión del tráfico y costes operacionales. Por otro lado, el terminal es capaz de elegir la red de acceso más adecuada, tal circunstancia obliga a desacoplar la tecnología de la infraestructura que presta el servicio para hacer ubicua la red de acceso del terminal móvil. La ubicuidad se consigue desarrollando algoritmos inteligentes capaces de asegurar un acceso sin cortes para cada uno de los diferentes tipos de servicio de usuario. El desacoplamiento de tecnologías obliga a buscar nuevas soluciones capaces de mantener la continuidad en la sesión durante el movimiento del terminal, y conseguir un uso eficiente de los recursos de red. En redes HetNets los recursos deben adaptarse a la movilidad del servicio en lugar de la tradicional movilidad del terminal en redes clásicas. Así mismo, hay que tener en cuenta que una mayor demanda de la tasa de bit por parte del servicio introduce nuevos problemas como la congestión de red, situación que obliga a readaptar las métricas de calidad de servicio (Quality of Service, QoS), a los requisitos del tráfico que demanda el servicio.

De ese modo, los servicios inalámbricos en redes HetNets introducen nuevos retos de QoS, servicios que generan un nuevo ecosistema que analiza el grupo de trabajo 3rd *Generation Partnership Project* (3GPP) [1]. El 3GPP, creado en diciembre de 1998, elige por consenso Long Term Evolution (LTE) como tecnología de banda ancha 4G, basada en el éxito de las redes de banda ancha 3G, tecnología para aprovisionar los futuros servicios móviles con tráfico asimétrico, tasa de bit variable, throughput elevado, provisión de QoS, pequeñas latencias, jitter moderado, baja perdida de paquetes, y baja señalización. LTE es capaz de proveer QoS tanto por usuario como por servicio, permitiendo desplegar servicios sensibles al retardo.

Para los operadores de red, LTE significa incrementar el rendimiento en términos de throughput. LTE se complementa como nexo de unión para soluciones HetNets, algunas de ellas utilizan espectro no licenciado con tecnología Wi-Fi 802.11n para formar el ecosistema que permite estructurar el enorme volumen de tráfico. El 3GPP especifica y define en sus especificaciones procedimientos de handover para integrar los diferentes accesos LTE HetNets.

La evolución de la tecnología inalámbrica desarrolla en paralelo y evoluciona la tecnología del backhaul, de ese modo el backhaul adquiere cierta relevancia debido a la capacidad de transmitir información paquetizada de modo eficiente a bajo coste. El backhaul agrega tráfico de varias redes, y de varios fabricantes de equipos. También es capaz de soportar QoS. Circunstancia que obliga a buscar nuevos modelos capaces de mejorar la excesiva señalización que surge debido a la convergencia de las redes de acceso.

El trabajo se centra en identificar nuevos modelos capaces de aprovisionar recursos para mantener la continuidad en la sesión en las redes de acceso con la menor intervención posible del usuario. La situación es especialmente delicada para servicios sensibles al retraso, servicios desplegados en infraestructuras de redes móviles de cuarta generación en entornos de movilidad IP. Entornos que surgen como consecuencia de la simplificación de la red móvil para hacerla transparente de la tecnología radio, tal circunstancia permite el acceso común a las diferentes tecnologías de acceso radio existentes. De ese modo, las arquitecturas adoptan estructuras planas debido a la consecuencia de tener una arquitectura de red orientada a paquetes.

La agregación de tráfico hacía el backbone común permite satisfacer las elevadas necesidades del cliente y del operador en términos de tasa de bit, circunstancia que obliga a buscar nuevos mecanismos de señalización y optimización de flujos, obliga también a mejorar la tecnología para ser capaz de entregar el servicio con el menor coste posible de señalización.

Índice

Lista de figuras

Lista de tablas

Acrónimos

3GPP	3rd Generation Partnership Project
6LSA	IPv6 Label Switching Architecture
AAA	Authentication, Authorization and Accounting
ABC	Always Best Connected
AH	Authenticator Header
ANDSF	Access Network Discovey and Selection Function
ARPU	Average Revenue Per User
BA	Binding Acknowledgement
BID	Binding Identifier
BU	Binding Update
CDMA	Code Division Multiple Access
CoA	Care-of-Address
CoS	Clases de Servicio
CQI	Channel Quality Indicator
DiffServ	Differentiated Service
D-GW	Distributed Gateway
DMM	Distributed Mobility Management
DSMIPv6	Dual Stack MIPv6
eNodeB	Evolved Base Stations
ePDG	Evolved Packet Data Gateway
ECRTP	Enhanced Compressed RTP
ESP	Encapsulating Security Payload
EPS	Envolved Packet System
E-UTRA	Evolved Universal Terrestrial Radio Access
FA	Foreign Agent
FDMA	Frequency Division Multiple Access
FIB	Forwarding Information Base
FMC	Fixed Mobile Convergence
GRE	Generic Routing Encapsulation
QCIF	Quarter Common Intermediate Format
HA	Home Agent
HetNet	Heterogeneous Cellular Networks
HMIPv6	Hierarchical Mobile IPv6
HNP	Home Network Prefix
HoA	Home of Address
HTTP	Hyper Text Transport Protocol
HSPA	High Speed Packet Access
GPS	Global Positioning System
GTPv2	Generic Packet Tunneling
IFOM	Flow mobility
IntServ	Integrated Services
ITU-T	International Telecommunication Union-Telecommunication
KPIs	Key Performace Indicators
L2TP	Layer 2 Tunneling Protocol
LAG	Local Access Gateway
LAN	Local Area Network
LDP	Label Distribution Protocol
LIPA	Local IP Access
LMD	Local Mobile Domain
LOS	Line-Of-Sight

LTE	Long Term Evolution
MAG	Mobile Access Gateway
M2N	Machine to Machine
MBMS	Multimedia Broadcast Multicast Service
MCoA	Multiple Care-of-Address
MPEG	Moving Pictures Experts Group
NEMO BS	Network mobility Basic support
NetMMN	Network-Based Localized Mobility Management
MIPv4	Mobile IPv4
MME	Mobility Management Entity
MN	Mobile Node
MNI	Mobile Node Identifier
MNPs	Mobile Network Prefixes
MOS	Mean Opinion Score
MTU	Maxium Transmission Unit
NOX	Network Operating System
MPLS	Multiprotocol Label Switching
NLOS	Non-Line-Of-Sight
MRSVP	Mobile Resource Reservation Protocol
OFDMA	Orthogonal Frequency Division Multiple Access
PCRF	Policy and Charging Rules Function
PBA	Proxy Binding Acknowledgement
PBU	Proxy Binding Update
P-GW	Packet Gateway
PMIPv6	Proxy Mobile IPv6
PPP	Point to Point Protocol
QoS	Quality of Service
QoE	Quality of Experience
RA	Router Advertisement
RAN	Radio Access Network
RRP	Return Routability Procedure
RTSP	Real Time Streaming Protocol
RTP	Real-time Transport Protocol
RTT	Round-Trip Packet Transmission
S2H	Service to Human
SDN	Software Defined Networking
S-GW	Serving gateway
SIPTO	Selected IP Traffic Offload
SLA	Service Level Agreement
SDF	Service Data Flows
TCAM	Ternary Content Addressable Memory
TCP	Transmission Control Protocol
TCRTP	Tunneling Multiplexed Compressed RTP
TDMA	Time Division Multiple Access
TEID	Túnel un Endpoint Identifier
TFT	Traffic Flow Agregates
TTL	Time-To-Live
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecomunicaction Services
VLC	Video LAN Client
VoD	Video on Demand
WCDMA	Wideband Code Division Multiple Access

1. Introducción y contexto

Los estándares proporcionados por los organismos para definir los sistemas inalámbricos actuales tienen como prioridad proporcionar la continuidad en la sesión IP estándares que se definen para ser capaces de desplegar los servicios de banda ancha. Servicios como *Service to Human* (S2H), o *Machine to Machine* (M2M) caracterizados por soportar altas tasas de bit y pequeño retardo. Es importante de ese modo establecer los requisitos que debe cumplir un determinado servicio, y a partir de ellos fijar los parámetros de servicio.

Requisitos como el s*eamless access*, acceso desde cualquier entorno o tecnología sin degradar la experiencia de usuario. El *Low handoff delay and loss rate*, métrica que minimiza la perdida de paquetes debido a los procedimientos handover. El m*ulti-service network*, aspecto esencial en redes de banda ancha, esto es la coexistencia de varios servicios. El b*roadband wireless access networks*, el ecosistema debe integrarse en la medida de lo posible con su entorno. El s*ecurity access and trafic*, el sistema debe integrar fuertes esquemas criptográficos, y gestionar la seguridad de forma adecuada. De ese modo, es necesario buscar y disponer modelos de QoS para modelar comportamientos de tráfico en los diferentes dominios de red para asegurar una gestión de recursos equitativa.

El factor más importante para servicios convergentes es la calidad extremo a extremo, el concepto de calidad encaja bien con métricas objetivas de QoS. A pesar de todo ello, el termino no es capaz optimizar el binomio recursos de red disponibles/percepción de usuario. Para ello se introduce un nuevo termino *Quality of Experience* (QoE), que tiene en cuenta los deseos de los consumidores.

Ello obliga a buscar nuevas aproximaciones para gestionar los recursos. Todas las mejoras actuales plantean una mejora del throughput que debe proporcionar la infraestructura, por lo tanto la capacidad de un sistema se define en base a la métrica de agregar tráfico en redes sectorizadas de forma que esta sea proporcional al número de usuarios existentes en tiempo según su tasa de bit, retardo o jitter.

Bajo ese punto de vista, incrementar el ancho de banda no es la mejor opción en términos de tiempo y presupuesto, y no debería ser la estrategia correcta para implementar una arquitectura basada en QoS, y QoE. Aunque el core de la estrategia es fácil de implementar, tecnologías como xDSL, LTE lo permiten. Desde el punto de vista estrictamente filosófico, el teorema "The Tragedy of The Commons" [2] describe como los recursos compartidos de un medio se agotan según el número de individuos que interactúan en su propio interés, a pesar de que su objetivo no es el agotamiento del recurso. Esto significa que el uso no administrado de los recursos conduce inevitablemente al agotamiento de los mismos. De ese modo, deben introducirse retos de calidad de servicio para manejar y optimizar los recursos en arquitecturas de Internet móvil. Arquitecturas que son capaces de proporcionar un servicio sin cortes, *seamless*, dentro de un contexto transparente en el acceso radio, acceso que proporcionan diferentes tecnologías de red, y donde la selección se realiza basándose en un perfil de servicio.

En la actualidad, existe cierta dificultad por parte de los operadores para proveer suficiente tasa de bit desde las estaciones base hasta el backbone, asegurando al mismo tiempo la movilidad dentro de un área extensa densamente poblada, áreas donde se despliegan las tradicionales arquitecturas de red monolíticas. De este modo, para conseguir altas tasas de bit se divide la celda en sucesivas celdas más pequeñas, cada celda contiene varias miniceldas con suficientes recursos cada una de ellas para proporcionar las altas tasas de bit demandadas por el servicio. En este nuevo escenario la movilidad y el roadming se posicionan como factores importantes para mantener la continuidad de la sesión cuando el nodo móvil (MN) realiza un handover. La continuidad de sesión debe ser transparente, que no se produzcan cortes en el servicio. Tradicionalmente, los handovers se basan en el indicador *Received Signal Strength* (RSS), las futuras arquitecturas deberán tener en cuenta otros factores como la QoE, las preferencias de

usuario, la movilidad, su localización, o los servicios utilizados por el MN, en el momento de decidir la posibilidad de un handover.

De ese modo, los nuevos estándares definidos por parte del 3GPP se desarrollan en ese sentido. Debido a las exigencias de un mercado en alza, el de la banda ancha. Estándares que son consensuadas por diferentes agentes y que adoptan tecnologías de reciente aparición. Posteriormente son llevados a industria debido a sus altos ratios de tasa de bit prometidos, junto su pequeño retardo, tecnologías como 802.11n, HSPA o LTE, aparentemente capaces de asegurar calidad de servicio. Todas ellas tienen en cuenta el factor QoS, como estrategia de futuro y de evolución. Actualmente las redes inalámbricas de banda ancha representan el lado más innovador de la tecnología, el concepto debe ser visto y analizado desde tres perspectivas diferentes. En primer lugar desde el punto de vista hardware, en segundo lugar desde el punto de vista software y finalmente desde el punto de vista de la tecnología de red.

En ese contexto se sitúa el trabajo, bajo el principio de red inalámbrica heterogénea y según el paradigma *Always Best Connected* (ABC). Visión que recoge la propuesta de la ITU-R M.1645 dentro del marco de trabajo de las futuras redes inteligentes. Donde los servicios necesariamente deben utilizar de forma eficiente los recursos de red disponibles, independientemente del fabricante de software o hardware [3]. De ese modo, se consigue cierta ubicuidad, y diferentes proveedores pueden operar sobre la misma infraestructura. Los servicios convergen hacia un entorno llamado *mobile broadband services*. En este contexto, la gestión del roadming entre tecnologías es fundamental. Arquitecturas donde es posible escalar el tráfico, abstrayendo los detalles de la infraestructura física hacía una infraestructura lógica, circunstancia que permite disminuir el tráfico de señalización y el overhead de red. Es posible soportar tareas concurrentes como encaminar, monitorizar, balancear, o inspeccionar el tráfico.

Tal circunstancia obliga al protocolo IPv4 a evolucionar de forma natural, básicamente debido a la escasez de direcciones posibles asignables, la escasa calidad de servicio, y la baja seguridad que proporciona el protocolo, hacia un nuevo protocolo llamado IPv6 capaz de direccionar un gran número de dispositivos, y soportar movilidad transparente en los nodos de red. En el caso del protocolo IPv4 la movilidad se plantea como una extensión del protocolo, en el protocolo IPv6 se mejora notablemente la gestión de la movilidad, la QoS y el tráfico multicast. Tal es el compromiso que el 3GPP define la entidad *Multimedia Broadcast Multicast Service* (MBMS) para sesiones punto a multipunto dentro de servicios portadores MBMS en arquitecturas de Internet móvil. La idea es ahorrar recursos radio compartiéndolos entre los usuarios de un grupo (TS 23.246. 2011).

Las especificaciones del protocolo IPv6 describen el comportamiento del protocolo ante flujos de tráfico. El protocolo se diseña, desde el principio pensando en la gestión de los flujos de tráfico, para ello introduce nuevos campos de cabecera capaces de identificar el flujo, y aplicar QoS. De ese modo, el 3GPP incorpora los nuevos desarrollos, y en paralelo desarrolla nuevos estándares con el objetivo de centralizar, y gestionar los nuevos escenarios de redes convergentes HetNets. Estándares que propone el IETF, y que son mantenidos por el 3GPP con el objetivo de conseguir bajos costes de capital para poder escalar el sistema bajo estructuras inteligentes basadas en la integración de los diferentes accesos de red.

Los estándares se desarrollan centrándose en la identificación del MN y del flujo de datos durante el tiempo del servicio. En una red móvil IP, la dirección IP de un nodo identifica inequívocamente el punto de conexión de dicho nodo. Así, un nodo móvil debe estar ubicado en una determinada red y caracterizado por un identificador de red a fin de poder recibir los paquetes que se le envían, de lo contrario los paquetes se pueden perder. Esto último se conoce como la gestión de la movilidad. La gestión de la movilidad esta formada por dos componentes principales, en primer lugar la gestión de la ubicación, permite descubrir el punto de conexión actual a la red de un nodo móvil para poder hacerle llegar información. En segundo lugar la gestión del movimiento, permite a una red mantener la conexión cuando un nodo móvil realiza

un movimiento y cambia su punto de conexión en la red. En este contexto, se exige conectividad total, también se exige QoS dentro de un entorno de roaming global con handovers verticales, transferir la sesión del usuario entre dos puntos de anclaje utilizando tecnologías diferentes. Existe en algunos casos solapamiento entre redes, también existe la posibilidad de elegir la red de acceso según criterios QoS.

En los nuevos ecosistemas que surgen, es necesario gestionar la autentificación de los dispositivos cada vez más heterogéneos que se incorporan en la red, hay que estabecer políticas de *Authentication, Authorization and Accounting* (AAA). Por lo tanto, se deben plantear diseños que consigan disminuir el tráfico de señalización para mejorar el aprovisionamiento de los recursos. Todas las innovaciones en arquitecturas de red avanzan en ese sentido, asignar recursos según necesidades y prioridades de servicio. Todas las arquitecturas están enlazadas por un único lenguaje IPv6, que diferencian los recursos físicos de los recursos lógicos, para minimizar de ese modo la gestión de la infraestructura y poder escalar la infraestructura.

El objetivo final de diseño de cualquier sistema busca mejorar su capacidad y aumentar los niveles de QoE. Tal es el compromiso que organismos como el 3GPP introducen cada vez más entidades de red responsables de señalizar recursos. Por ejemplo, el 3GPP define una entidad *Access Network Discovey and Selection Function* (ANDSF) capaz de proporcionar información a los dispositivos para que puedan decidir el mejor acceso a la red según preferencias de usuario previamente definidas. En realidad actualmente el ANDSF se utiliza para interconectar redes 3GPP con redes non-3GPP y establecer handovers predictivos. El documento TR 22.912 del 3GPP especifica los procedimientos para seleccionar el mejor acceso, siempre y como fin último el de proporcionar un nivel adecuado de servicio.

1.1. Ubicación del trabajo

El despliegue de servicios de banda ancha en redes convergentes no está exento de desafíos. Redes que surgen como consecuencia del nuevo paradigma surgido dentro de un contexto de arquitectura de red centrada en el usuario. Paradigma que aumenta la complejidad en las interacciones entre los protocolos y los sistemas, llegando a tener en algunos casos comportamientos incorrectos y agujeros de seguridad.

En este nuevo contexto se posicionan como factores relevantes, la gestión del estado *seamless*, la continuidad de la sesión IP durante el movimiento del terminal en las distintas redes de acceso y la provisión QoS en el flujo que caracteriza el servicio. La continuidad de la sesión IP significa conseguir pequeñas latencias, jitter moderado, y un número reducido de retransmisiones de paquetes durante el movimiento del MN entre las distintas redes de acceso. Con esa visión se están diseñando los nuevos protocolos de movilidad, en especial *Proxy Mobile IPv6* (PMIPv6), protocolos capaces de asegurar una gestión transparente de la movilidad en la capa de red y de maximizar en la medida de lo posible la QoS durante el movimiento del terminal. El protocolo PMIP es el elegido por el 3GPP para ser el candidato que suministre los nuevos servicios en entornos de movilidad, con requisitos de tasa de bit y sensibles al tiempo.

Los esfuerzos actuales de los organismos de estandarización se focalizan en ese sentido, como estructurar el enorme volumen de tráfico. La situación está amplificada en las redes de acceso, debido a la naturaleza del espectro, caro y limitado por naturaleza, circunstancia que obliga a buscar nuevas soluciones. De ese modo, las *small-cells*, se convierten en agentes importantes dentro de la estrategia del operador para ofrecer servicios de banda ancha, debido a su capacidad de integrarse en las redes de acceso.

La situación esta favorecida por la introducción del protocolo IPv6 tanto en el core de red, como en las redes de acceso, tal circunstancia cambia el paradigma de una estrategia de gestión de QoS. El nuevo contexto que surge despierta gran interés en la comunidad científica, y donde se enmarca este trabajo. Un análisis en ese contexto demuestra que es posible encontrar nuevas

propuestas capaces de optimizar la latencia del servicio, y mejorar la capacidad del sistema en términos de throughput para desplegar servicios de banda ancha con la QoE esperada por el usuario.

El trabajo presenta el problema de la movilidad IP que surge como consecuencia del movimiento del terminal por las diferentes áreas de cobertura que proporciona LTE, red orientada a paquetes. Como consecuencia, la arquitectura de LTE se convierte en una arquitectura plana, circunstancia que trae implícito el problema de cómo estructurar el enorme volumen de tráfico que agregan las redes de acceso.

Por lo tanto, una gestión completa de la movilidad debe contemplarse desde la perspectiva de la optimización de los protocolos en los mecanismos de *handover*. Mecanismos orientados hacía la continuidad del servicio en entornos intra-area, mismo dominio tecnológico o inter-area, diferentes dominios tecnológicos, handovers verticales entre LTE y WiFi.

1.2. Objetivos del trabajo

La principal aportación del trabajo consiste en analizar y proponer posibles nuevos métodos que puedan ser utilizados para ofrecer servicios extremo a extremo (E2E) con QoS en redes HetNets. En este sentido, el crecimiento de las redes de banda ancha permite definir nuevas arquitecturas de servicio, con el objetivo de integrar las principales funciones de las redes de acceso móvil de tercera generación y aquellas consideradas de cuarta generación.

El análisis se desarrolla desde la consideración inicial del IETF donde la inteligencia de red se ubica en los MN, y concluye el planteamiento proponiendo mecanismos en redes móviles para ubicar la inteligencia de red, en la red misma, a través de nodos de servicios. El trabajo concluye que de ese modo se puede mejorar el diseño de redes 4G con el objetivo de favorecer la movilidad, mejorar la infraestructura de red y la seguridad, para ello es necesario que la inteligencia de la red sea proporcionada a través de nodos de red específicos.

De ese modo, el trabajo analiza los mecanismos propuestos por organismos como el 3GPP y el IETF para estructurar el tráfico y reducir la latencia del servicio, finalmente el trabajo propone nuevos mecanismos para optimizar el tráfico mediante flujos con optimización de cabecera en la parte fija de la red.

Actualmente el problema está amplificado debido a la demanda en la tasa de bit por parte de los usuarios móviles. Usuarios que presionan a los operadores para mejorar los servicios *seamless*, de ese modo los operadores se ven obligados a optimizar los protocolos, debido a que no se diseñaron para proporcionar continuidad en la sesión. Tampoco se diseñaron para transportar servicios en tiempo real sensibles al retardo.

La principal diferencia en el acceso a red respecto al pasado tiene lugar en la selección del acceso a red, cada vez más heterogéneo. Tal circunstancia está favorecida por la configuración multi-interfaz del terminal, en la que los *handovers* pueden realizarse a partir de un perfil de servicio independientemente de la tecnología o dominio administrado.

Por otro lado, la complejidad actual en la gestión de redes hace difícil aplicar, introducir y mejorar los mecanismos de gestión de red, tales como establecer políticas de seguridad, QoS, y las reglas de acceso. La situación es extremadamente sensible desde el punto de vista de los acuerdos *Service Level Agreement* (SLA). Por lo tanto, el despliegue de nuevos servicios de banda ancha E2E debe contemplar un cambio de paradigma en la gestión del tráfico.

Por ejemplo, para un servicio de video en tiempo real, el tiempo de interrupción requerido en el proceso de un *handover* entre LTE y UMTS debe ser inferior a 300 msg.

Por otro lado, para un servicio VoIP con un retardo menor de 300 ms, el *handover* tendrá éxito en el 99% de los casos. Hay que tener en cuenta que el tiempo para reenviar paquetes *Round-Trip Packet Transmission* (RTT) en UMTS está entre 150ms, y 200ms. En LTE está en 15ms,

y en HSPA+ entorno a 50ms y 100ms. Son valores típicos que pueden encontrarse en el forum 3GPP (TS 36.133).

El análisis del problema desde la perspectiva de la selección de red permite optimizar los recursos a partir de algoritmos dinámicos. Los recursos se asignan según las necesidades de usuario y de servicio. Los terminales toman las decisiones para traspasar el flujo de tráfico desde una red de acceso a otra. El mecanismo es útil en sistemas autónomos y redes convergentes.

De ese modo, los servicios ofrecidos por los operadores pueden orientarse hacia un *Internet offload*, esto significa que el tráfico de datos y de señalización debe tener el menor impacto posible en la infraestructura. El enfoque consigue maximizar los ingresos de los operadores y minimizar el coste del servicio entregado. Aunque el concepto introduce una enorme complejidad en las redes de acceso, *handovers* y en el manejo de la QoS. Hay que tener en cuenta que el diseño de arquitecturas de red impacta directamente en los servicios, especialmente aquellos que son sensibles al retardo, como VoD en tiempo real. Situación especialmente dedicada en entornos heterogéneos.

El resto del trabajo se estructura del siguiente modo, en el apartado segundo se analiza el estado del arte de la problemática actual en entornos HetNet. En el apartado 3 se describen y analizan los mecanismos actuales de gestión de la movilidad junto a las diferentes variantes del protocolo MIP orientadas a la disminución de latencias que consiguen aumentar el throughput, se analizan propuestas 3GPP y IETF que mejoran la latencia. En el apartado 4 se proponen nuevos mecanismos para optimizar la señalización de los flujos de tráfico y reducir de ese modo los tiempos de traspaso en la continuidad del servicio. Se plantean como posibles soluciones para que los operadores de red las puedan adoptar e introducir en su *roadmap* con el objetivo de proveer servicios *seamless* sobre LTE, finalmente se plantean las conclusiones y las líneas futuras.

2. Situación tecnológica

La imparcialidad en la asignación de los recursos es uno de los aspectos que caracterizan las arquitecturas de red tradicionales. Los usuarios compiten entre sí por el recurso, en contraposición de las arquitecturas de nueva generación capaces de proporcionar los mecanismos y los recursos necesarios para desplegar servicios con requisitos QoS.

El diseño de arquitecturas tradicionales se focaliza en la capacidad del terminal y la duración sus baterías, posteriormente las arquitecturas de nueva generación añaden la complejidad de las tecnologías de acceso (*Radio Access Network* (RAN)). Tecnologías como HSPA, o LTE capaces de proporcionar altas tasas de bit, y una reducida latencia. Tal circunstancia produce un cambio significativo, consigue aumentar las posibilidades de éxito para entregar servicios con alta QoE, situación en la que nos encontramos actualmente.

Detrás del grupo de trabajo 3GPP subyace un conjunto de recomendaciones que especifican la evolución del núcleo de red de los diferentes estándares propuestos por el organismo, como GSM/WCDMA/HSPA/LTE/LTE-A. De ese modo, la arquitectura LTE-A integra todas las tecnologías siendo una tecnología totalmente orientada a conmutación de paquetes.

LTE no soporta nativamente servicios de conmutación de circuitos, por el contrario el *Evolved Packet System* (EPS), formado por el *Evolved UMTS Terrestrial Radio Access Network* (E-UTRAN) formado por el eHB, y el EPS es el *Evolved Packet Core* (EPC) núcleo IP E2E, que es totalmente IP con protocolos estandarizados por el IETF y capaces de agregar tráfico tanto para accesos fijos como móviles. El EPC consigue hacer menos dependiente las tecnologías RAN. Aunque tal circunstancia introduce un nuevo problema, mantener la continuidad del servicio cuando los usuarios se mueven entre diferentes coberturas proporcionadas por LTE. El EPC es capaz de interoperar con diferentes tecnologías de acceso 2G, 3G, 4G, y WiFi. El EPC

tiene muchos principios comunes con IMS, como la capacidad de soportar QoS, o la autentificación de usuario. Sin embargo algo que diferencia al EPC es el concepto de continuidad en el servicio en la capa de red para proporcionar una conexión sin cortes.

El EPC lleva a cabo la unión de diferentes tipos de tráfico fijo o móvil hacia el *Packet Gateway* (P-GW). Gestiona la movilidad entre estaciones base y redes de acceso a través del *serving gateway* (S-GW). Gestiona la congestión y provee QoS para aplicaciones sensibles al retardo. Lleva a cabo un rol crucial en redes HetNets la autentificación, el accounting y la autorización.

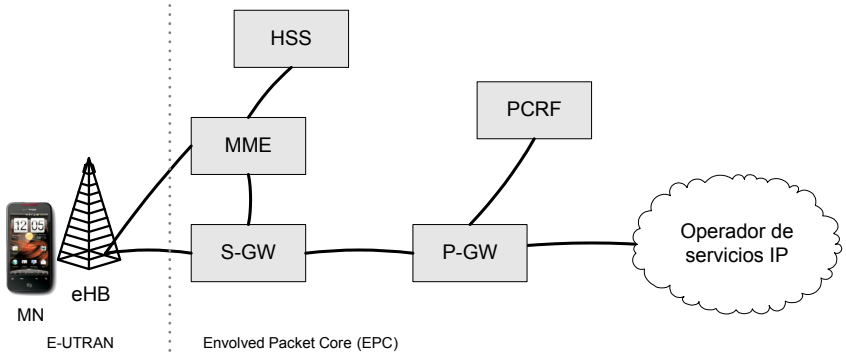

Figura 1. Arquitectura típica en LTE

Por otro lado al sistema extremo a extremo, la combinación de la red de acceso E-UTRAN y la red troncal EPC, se le denomina Envolved Packet System (EPS) Release 8 del 2010, algunos autores asocian EPS a LTE, convención que sigue el trabajo. El sistema soporta accesos radio 3GPP y accesos non-3GPP. El EPS soporta handovers entre accesos 3GPP y non-3GPP, las nuevas propuestas mejoran la seguridad y soportan la movilidad basada en PMIP[3], protocolo catalogado de red que consigue que el terminal no este involucrado en la señalización. En este contexto diversos autores plantean sus propuestas y recomendaciones orientadas a la movilidad proporcionada por las múltiples tecnologías de acceso radio. Permitiendo que usuarios puedan intercambiar una conexión, por ejemplo entre LTE y WLAN o entre LTE y HSPA+.

El 3GPP en sus especificaciones establece los requisitos que deben cumplir las redes basadas en el protocolo IP. Las especificaciones proporcionan diseños de redes *seamless mobility* modo de referirse a la continuidad del servicio cuando se realizan handovers entre varios tipos de acceso. El grupo de trabajo diseña recomendaciones siempre pensando en que la infraestructura pueda escalar y que esta sea mantenible bajo un eficiente uso del ancho de banda. El grupo de trabajo minimiza el número de elementos de red, para tener de ese modo estructuras menos jerárquicas, estructuras capaces de disminuir la latencia del servicio, utilizando para ello una menor señalización y consiguiendo una menor perdida de paquetes.

De ese modo el EPC se convierte en una pieza fundamental en las redes de banda ancha, sin el EPC ninguna tecnología de acceso RAN o servicio proporcionado por la infraestructura puede alcanzar todo su potencial.

En sistemas 3GPP y por la importancia que adquiere, el tráfico dentro del EPC se encapsula a través del protocolo Generic Packet Tunneling (GTPv2) sobre IP y UDP, túneles encargados de enlazar un nodo de red hacia el gateway (P-GW), túneles por los que se encaminan los diferentes flujos de datos pertenecientes al MN, cada MN crea un único túnel que comparte el uplink con el downlink. La gestión de los túneles está centralizada y gestionada por el EPC, mientras que la gestión de rutas es distribuida, esto significa que cuando cambia una ruta los elementos de red deben reconfigurarse, circunstancia que produce inestabilidad en el sistema. Cada túnel está identificado en cada nodo a través de *Túnel un Endpoint Identifier* (TEID), una dirección IP y un puerto UDP.

El uso intensivo de túneles fue necesario en los primeros diseños de redes celulares para transportar los servicios sobre la infraestructura, el tráfico debe transportarse de algún modo, y

la encapsulación del tráfico permite abstraer a los elementos de red. El túnel no es malo en si aunque crea overhead de red, y una masiva concentración de túneles en el P-GW puede ser perjudicial, debido a un consumo elevado de recursos. Muchos túneles gestionados por el P-GW pueden no ser adecuados para mantener la sesión del MN en las nuevas arquitecturas de movilidad, en las que existe una alta probabilidad de realizar handovers.

Desde ese punto de vista es posible analizar nuevos mecanismos de movilidad [4] [5]. Debido fundamentalmente a la pequeña cobertura de las *small-cells*, situación que favorece la búsqueda de nuevos mecanismos para mantener la continuidad en la sesión con la macrocelda. En principio las *small-cells* actúan como cualquier otra estación base utilizando los mismos procedimientos de *handover*. Aunque la continuidad en la sesión debe ser analizada desde otra visión, actualmente es donde el 3GPP pone mucha atención, desarrollando nuevos procedimientos capaces de mantener la sesión entre las *small-cells* y tecnologías non-3GPP como WiFi.

El backhaul comprende los enlaces desde la red de acceso hasta el backbone de red. En 3G el backhaul transporta los datos de la celda hacia el backbone de red, a través de sistemas inalámbricos o fibra óptica. En LTE el transporte se puede realizar también por IP/Ethernet, de ese modo es posible conseguir mayor eficiencia que cuando se utiliza la interfaz radio.

Por lo tanto, la eficiencia en el diseño de la arquitectura depende tanto del dimensionado de la celda, como la optimización del backhaul. También es posible diseñar estrategias para encapsular el tráfico mediante un túnel para cada servicio, o acceder a servidores diferentes que proporcionar un mismo servicio y de ese modo aumentar los niveles de QoS.

La posibilidad de dotar al backhaul de capacidades IP permite agregar el tráfico generado en el interior de una small-cell, liberando recursos de la macrocelda para otros usuarios móviles. El enfoque debe contemplar estrategias QoS debido a que de modo natural IP no soporta QoS.

El esfuerzo para optimizar el backhaul, obliga a añadir mecanismos capaces de transformar las redes de acceso en arquitecturas de red orientadas a servicios. Como consecuencia, nace la tecnología *Multiprotocol Label Switching* (MPLS), tecnología que consigue de modo natural unificar los diferentes tipos de tráfico en el backhaul utilizando para ello la conmutación de etiquetas. El mecanismo optimiza el tradicional proceso de encaminamiento de paquetes. El tráfico se clasifica, para enviarse a los routers de entrada en la red del proveedor, donde el tráfico se encamina hacia los routers de salida, para agregarse finalmente al backbone de red. De ese modo, los proveedores despliegan sus servicios utilizando MPLS. La tecnología consigue asegurar la disponibilidad E2E del servicio con garantías de QoS.

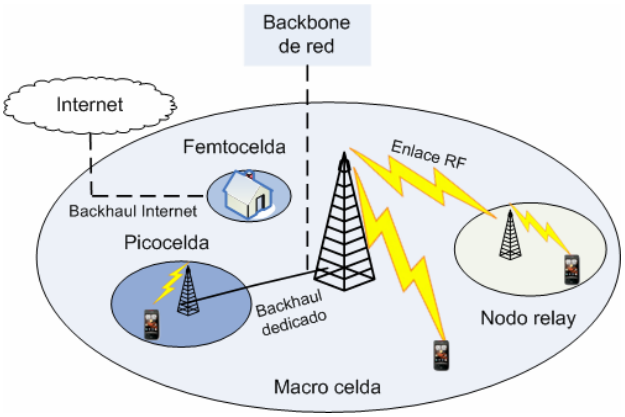

Figura 2. Red heterogénea (HetNet)

Los proveedores de servicios, impulsados por la reciente demanda de ancho de banda y la necesidad del crecimiento *Average Revenue Per User* (ARPU), ingresos medios por usuario, migran rápidamente sus infraestructuras hacía la próxima generación de tecnologías móviles, como LTE y adoptan mecanismos MPLS para implementar el backhaul. LTE ofrece un coste

menor en términos de ancho de banda, mejora la prestación de servicios, permite un uso flexible de las bandas de frecuencias y simplifica notablemente la arquitectura de red. El ethernet backhaul ofrece una mayor capacidad de conexión a un menor coste por bit, lo que se traduce a mayores beneficios para el proveedor de servicios [6]. De ese modo, para los operadores una estrategia para optimizar el backhaul lleva implícita una disminución de la congestión. La congestión produce retardo y jitter, circunstancia no deseable para la QoE.

2.1. Arquitectura WCDMA y su evolución hacia HSPA/LTE/LTE-A

Universal Mobile Telecomunicaction Services (UMTS) se diseña desde el principio para ser una red flexible en la entrega de tráfico, de cualquier tipo de tráfico. Así pues, las tecnologías *Wideband Code Division Multiple Access* (WCDMA) y *High Speed Packet Access* (HSPA) son capaces de albergar los nuevos servicios, bajo las siguientes características:

- Se disminuye la complejidad de la arquitectura.
- Proporcionan altas tasas de bit, teóricamente hasta 2 Mbps en la Release 99, hasta 14.4 Mbps en la Release 5 del 3GPP, llegando hasta 28,8 Mbps en la Release 7 siendo actualmente el bit rate de 100 Mbps en la Release 11 del 3GPP. En el momento de escribir este trabajo se está trabajando en la Release 12.
- Se imponen como requisitos de diseño niveles de retraso en paquetes de ida y vuelta pequeños. Se fijan por debajo de 100ms en la Release 5, y debajo de 50ms en la Release 6.
- Se introduce el concepto de conexión sin cortes en un servicio de datos.
- Existe la posibilidad de aplicar QoS en flujos de datos, tratamiento diferenciado a los paquetes transmitidos para que el paquete llegue a su destino con el menor retardo posible.
- Es posible simultanear diferentes tipos de servicios de datos, por ejemplo servicios sensensibles al retardo, servicios interactivos, o servicios clasicos *best effort*.

En la Release 8 congelada en diciembre del 2010 (TR-36.913), la *International Telecommunications Union - Radiocommunications* (ITU-R) define los requisitos para el *International Mobile Telephony - Advanced* (IMT-A), con el objetivo de estandarizar la cuarta generación de redes inalámbricas (4G). Tecnologías que prometen gran caudal de tráfico tanto en el enlace ascendente como en el enlace ascendente. En marzo del 2011, la Release 10 se establece como candidata por parte del grupo 3GPP, para cumplir los requisitos definidos en el IMT-A, entre los requisitos más importantes están la gestión de la QoS, *Extended-Multiple-Input-Multiple- Output* (E-MIMO), *Coordinated Multi-point Transmission* (CoMP), *Relaying, y LTE Self-Optimizing Networks* (SON). De ese modo, la Release 10 se transforma en *Long Term Evolution - Advanced* (LTE-A). El calendario del 3GPP está alineado con el calendario de la ITU, sin embargo, la visión dentro del 3GPP es que LTE-A no debe limitarse a cumplir los requerimientos de IMT-A, sino que debe perseguir objetivos más ambiciosos. Las características principales de LTE-A son:

• Soportar varias redes de acceso y la posibilidad que la selección de la red de acceso se configure a partir de las políticas de operador, preferencias del usuario y las condiciones de la red de acceso.

- Flexibilidad en el uso del expectro radioeléctrico. Bit rate de 500Mps con una eficiencia espectral de 2.5 bps/Hz, en el enlace ascendente. Y tasas de 1Gbps con una eficiencia espectral de 5 bps/Hz en el enlace descendente, cuando el MN está en reposo. Para un MN en movimiento, se permite tasas de 100 Mpbs, dentro de un ancho de banda de 20MHz.
- Conectividad sin cortes, y posibilidad de *handovers* entre redes heterogéneas para mejorar la cobertura haciendo un uso razonable de las baterías del terminal.

- Latencia mejorada hasta 5 ms en el plano de usuario y 50 ms en el plano de control. Es posible tener sesiones interactivas en tiempo real, debido al menor descarte de paquetes como consecuencia de la latencia.

- Se introducen mecanismos para soportar movilidad entre dominios y entre redes. Posibilidad de seleccionar el mejor acceso de red según perfiles de usuario.

- Finalmente, los requisitos más importantes introducidos en las últimas Releases se refieren a la seguridad, y la gestión QoS. QoS integrada entre varias tecnologías y dominios *seamlessly* capaces de entregar contenido al usuario con alta QoE.

Parámetro	**FDMA**	**TDMA**	**CDMA**	**OFDMA**
Throughput	Bajo	Bajo	Moderado	Alto
Técnica de acceso	Selección de frecuencias	Slot	Espectro ensanchado	Selección de frecuencias
Latencia	Bajo	Incrementa con el número de usuarios	Bajo	Constante, y bajo
Ancho de banda por canal	0.02 MHz	0,02 MHz	1,25 MHz	1,5 – 2,5, 5 – 10 – 15 - 20 MHz
Ancho de banda dinámico	No	Sí	Sí	Sí
Gestión de potencia	No	No	Sí	Sí
B.E.R.	Alto	Alto	Bajo	Bajo
ISI	Bajo	Bajo	Mitigado	No existe

Tabla 1. Comparativa entre las diferentes técnicas de acceso

2.2. Arquitectura EPS

En este apartado se analiza la arquitectura EPS debido a la importancia que tiene para las conclusiones del trabajo. La arquitectura EPS esta formada por diferentes dominios lógicos de acceso, dominios RAN conectados al EPC de modo coordinado para proporcionar las funcionalidades requeridas por parte de la infraestructura al usuario. La figura 3 ilustra una implementación funcional del modelo a partir de las especificaciones propuestas por el 3GPP. El dominio de señalización opera en el núcleo de red, dominio que provee información de control, dominio responsable de señalizar la movilidad y de coordinar al resto de dominios funcionales. Así mismo el dominio de señalización proporciona niveles de QoS E2E a través de un servicio de portadoras en dominios RAN.

El servicio de portadoras define la granularidad de la QoS a partir de criterios técnicos definidos por el operador con la posibilidad de encaminar flujos de datos. El servicio portador separa el tráfico para proveer diferentes tratamientos según requisitos de QoS. El sistema señaliza los recursos que deben reservarse para el servicio portador, antes que los flujos de paquetes sean mapeados al servicio portador.

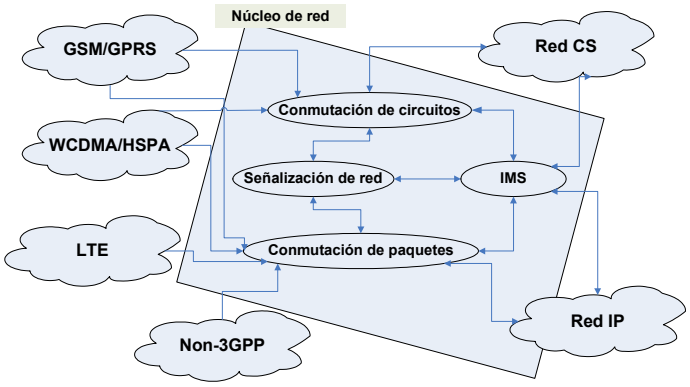

Figura 3. Dominios funcionales (TS 23.402-v12.0)

Todos los flujos de paquetes mapeados a un portador reciben el mismo tratamiento, como la política de scheduling, la gestión de colas y la configuración. En el caso de que existan flujos

diferentes, en cuanto a requisitos para un determinado servicio se utilizan servicios portadores diferentes. Un portador es una combinación de los parámetros QoS y la IP del terminal. La relación entre el servicio portador EPS y la QoS se establece mediante clases de servicio donde a cada portador se le asigna solamente una clase de servicio. Cada clase está identificada por la red del flujo de paquetes asociado a la IP del MN. Todas las comunicaciones IP extremo a extremo proveen de una cabecera túnel, la cabecera contiene el identificador de portador, de ese modo los nodos pueden asociar el paquete con los correspondientes parámetros QoS. También contiene un código de diferenciación del servicio. La relación entre el servicio portador EPS y la QoS se establece mediante clases de servicio donde a cada portador se le asigna solamente una clase identificada por la red al flujo de paquetes asociado a la IP del terminal.

La red de acceso es responsable de mantener la transmisión radio con los equipos de usuario para proporcionar la conectividad necesaria entre los MNs y los equipos de la red troncal. Los servicios de transmisión ofrecidos por la red de acceso para transportar la información de los equipos de usuario, tanto información de datos como de señalización, hacia/desde la red troncal son los servicios portadores. La red de acceso es la responsable de gestionar el buen uso de los recursos radio disponibles para la provisión de servicios portadores de forma eficiente.

Los diferentes dominios de acceso difieren en ancho de banda, latencia o coste, de ese modo una estrategia QoS no puede basarse solamente en métricas cuantitativas, como jitter, perdida de paquetes, ancho de banda, o el retardo. Es necesario analizar de modo intrínseco la QoE para tomar una decisión que seleccione el mejor acceso a red. De ahí la dificultad en la implementación de mecanismos de selección de acceso basados en estrategias QoS, mecanismos que toman especial relevancia en los modelos de red Wireless actuales centrados en una perspectiva orientada al usuario. En este sentido se definen los Key Performace Indicators (KPIs), conjunto de métricas recogidas en el TR 32.814 con el objetivo de evaluar el rendimiento en redes Wireless para proporcionar aceptable QoE. De ese modo, un indicador importante es el coste de la red de acceso, por lo tanto el acceso a red esta influenciado por factores como criterios los de red, los criterios de servicio, y los criterios de usuario.

La red troncal es parte del sistema encargado del control de acceso, la autenticación de los usuarios en el sistema, la gestión de la movilidad de los usuarios, la gestión de las sesiones de datos que transportan la información de los usuarios, y los mecanismos de interconexión con otras redes. Es muy importante en la red troncal disminuir, en la medida de lo posible, el tráfico de señalización, y que el aprovisionamiento de recursos se realice de un modo equitativo.

La figura 4 ilustra una arquitectura EPC, representa las diferentes redes de acceso, redes como 3G, 4G, y non-3GPP. El trabajo se centra en el core de la arquitectura LTE, donde sus entidades principales son el *Evolved Base Stations* (eNodeB), resposable de la gestión de recursos radio. Cuando un paquete llega al eNB comprime la cabecera y encripta los datos, la entidad responsable de añadir la cabecera GTP en el plano de usuario, y enviar el paquete al S-GW, también interroga al MME para mantener el sistema portador. Como es el único responsable en la parte radio lleva a cabo la gestion de la QoS en EUTRAN. *El Mobility Management Entity* (MME) que lleva a cabo la señalización entre el eNodeB y el SGW que es el punto de anclaje local móvil. El MME también provee el identificador de red al MN, que coordinándose con el eHB lleva a cabo la señalización salto a salto. El MME gestiona el establecimiento, y el cierre de la sesión además de la reconfiguración de la sesión del MN. Maneja la movilidad, actualiza la posición del MN, lleva a cabo el aviso de llamada, y gestiona los handovers. El P-GW es la pasarela que interconecta redes externas, esto significa que la información de un usuario conectado al sistema LTE pueda ser visible desde una red externa. Por ejemplo, en una sesión VoIP el P-GW envía información de QoS, y la tupla que caracteriza el tráfico al S-GW. El S-GW envía la información al MME para que aprovisione los recursos en la estación base. En el momento del *handover*, el MN notifica el cambio al MME que libera, y aprovisiona recursos desde la antigua hasta la nueva estación base.

Por lo tanto los paquetes IP generados por el usuario se inyectan en la red externa a través de la pasarela y viceversa. Todo el tráfico IP dirigido a un terminal LTE proveniente de la red externa va a ser encaminado hasta el P-GW a través de las políticas contenidas en el *Policy and Charging Rules Function* (PCRF), elemento que activa, o desactiva ciertos parámetros QoS, y lleva a cabo el control de la tarificación de usuario. Entidades que todas ellas constituyen los elementos que aprovisionan un servicio, y proporcionan conectividad IP entre equipos de usuario conectados a través de la E-UTRAN.

La pasarela P-GW actúa de punto de anclaje para la gestión de movilidad entre LTE y redes non-3GPP. La pasarela asume funciones de *Home Agent* (HA) funciones capaces de proporcionar continuidad en el servicio en caso de utilizar el protocolo MIPv4. Gestiona la movilidad entre una red LTE y una red WiFi. Además en MIPv4, la pasarela incluye soporte *Dual Stack* MIPv6 (DSMIPv6) y PMIPv6. La pasarela asigna una dirección IP al terminal de red, finalmente la pasarela P-GW puede ser vista como un router IP convencional.

La separación entre la red de acceso y la red troncal dota de flexibilidad al sistema de cara a soportar un proceso evolutivo en donde se pueda añadir o sustituir las diferentes partes de la red para escalar el sistema. También permite que los servicios puedan ser entregados a diferentes redes. Las entidades de red que se describen en la arquitectura de la red troncal son entidades funcionales, una entidad de red en 3GPP se concibe como una entidad lógica que cubre una funcionalidad que la funcionalidad está perfectamente definida. Hay que tener en cuenta que diferentes entidades funcionales pueden residir en el mismo equipo físico.

En este párrafo se enumeran los interfaces más representativos de la arquitectura. La interfaz S1-U utiliza el protocolo GTP-U para transportar los datos, la interfaz S3 utiliza el protocolo GTP [7], la interfaz S4 esta basada en el protocolo GTP, la interfaz S5 utiliza los protocolos GTP, y PMIP como modos de transporte (TS 23.402), la interfaz S8 utiliza los protocolo GTP, y PMIP (TS 23.402). El interfaz S9 provee transferencia de parámetros QoS desde la red actual hacia la visitada (H-PCRF, V-PCRF) [8].

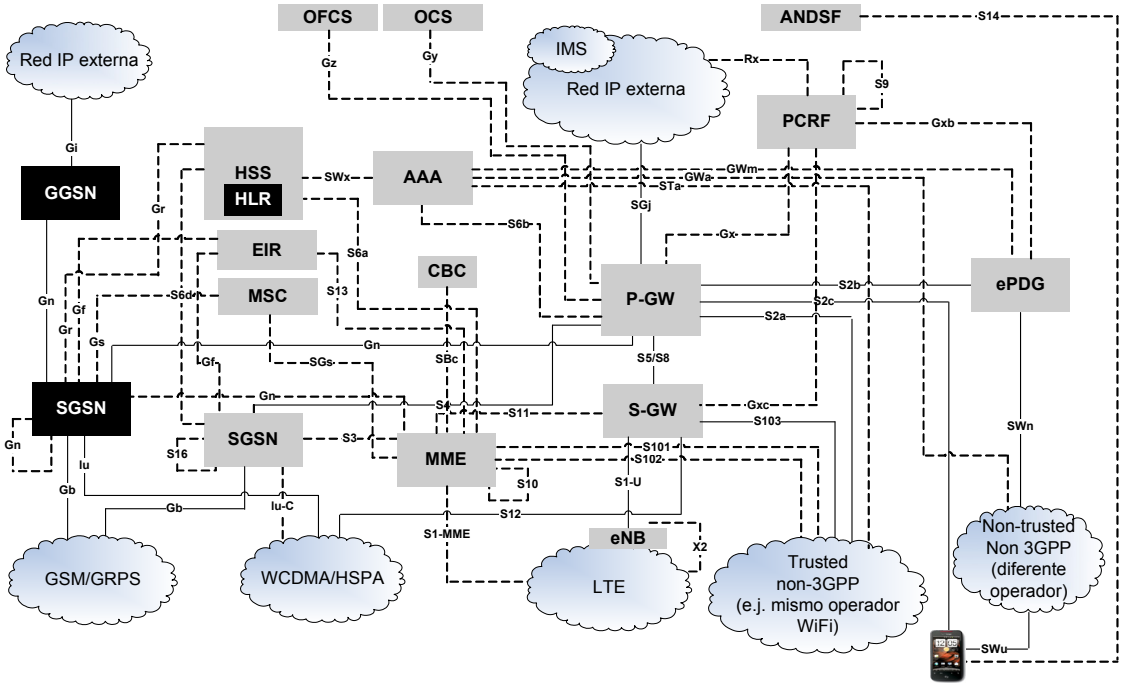

Figura 4. Arquitectura lógica desarrollada para el EPS (3GPP TS 29.274-v12)

2.3. Movilidad en redes móviles orientada a la QoS

Cuando el usuario del terminal móvil cambia de celda anuncia el cambio mediante anuncios de ruta en capa 3 o disparadores en capa 2. El procedimiento transfiere el control de la

comunicación desde la estación base inicial a una nueva estación base en la nueva celda. Durante el intervalo de tiempo que dura este proceso los paquetes destinados al MN y aquellos transmitidos por el MN no pueden ser enrutados. Esto provoca pausas o latencias en el servicio, también provoca cierta perdida de paquetes.

Para efectuar un traspaso se deben tener en cuenta diversos factores como los criterios de decisión del MN, la recogida de información a partir de métricas, estimar el peso para cada una de las métricas recogidas, y finalmente explorar todas las alternativas posibles en los nuevos accesos de red para decidir si el traspaso es necesario. De ese modo, diferentes servicios tienen estrategias de *handover* diferentes. Por ejemplo, en entornos de alta movilidad la degradación de la señal será la decisión más importante para el *handover*, en otros escenarios el factor será el coste, la perdida de paquetes, el número de paquetes retransmitidos, o la tasa de bit.

Las estructuras celulares jerárquicas llevan a cabo una reducida cantidad de traspasos, esto se consigue mediante una apropiada asociación de los usuarios en las distintas capas de la celda según sea su movilidad. Así por ejemplo, es mejor soportar una comunicación desde un vehículo en movimiento a través de macroceldas que microceldas, ya que en este último caso se deben realizar handovers muy frecuentemente. Se trata de evitar que los usuarios que presentan una elevada velocidad vean reducido su factor de movilidad mediante la asignación de una celda que presente un radio mayor.

De este modo, las nuevas arquitecturas de *small-cells* dentro del marco de los sistemas 4G permiten aumentar las velocidades de transmisión de pico, sin malgastar la capacidad de red para penetrar en interiores, donde se genera gran parte del tráfico a cursar. Otra ventaja que aportan las *small-cells* es que pueden hacer uso de las bandas de frecuencia más altas al tener asociadas coberturas limitadas.

Por lo tanto, un *handover* implica una gestión en la capa de red (L3), para cambiar la información de ruta y la dirección IP. Además de una gestión en la capa Wireless para cambiar las conexiones L2. En este sentido, la tecnología MIP provee una capa común a todas las tecnologías que proporciona movilidad y mantiene la continuidad en la sesión.

En este sentido, en el diseño de EPS se separó de modo explicito los accesos de red de los servicios ofrecidos por el proveedor de servicios, con el objetivo de conseguir una red de acceso transparente al core IP. Por otro lado, la interrupción del servicio durante el traspaso en capa 3 es un factor crucial para servicios en tiempo real.

Las métricas QoS en redes Wireless basadas en IP se definen por la perdida de paquetes, el retraso en el *handover*, y el *overhead* en la señalización del servicio. Todas las propuestas que se plantean buscan minimizar esas métricas y la actualización de rutas dinámicamente cuando el MN cambia su punto de anclaje en la red. Cada movimiento que realiza el MN requiere de señalización. El cambio de ubicación del MN puede producirse desde una topología física hacia una móvil o viceversa. En ese sentido, el IETF provee y especifica números protocolos capaces de señalizar los *handovers* que realiza el MN. Emerge entonces la gestión de movilidad en capa de red, capa responsable de mantener la continuidad de la sesión de los diferentes agentes móviles. La capa oculta el movimiento del MN a protocolos de capas superiores, y mantiene la continuidad en la sesión.

La latencia en el *handover* se define como el tiempo transcurrido desde que se detecta la petición de *handover* en capa 2, hasta que el MN recibe el primer paquete en la nueva posición de red. En entornos inalámbricos, y debido a la naturaleza del radio canal, las latencias deben gestionarse, buscando mecanismos capaces de mitigarla. Todas las mejoras se orientan en ese sentido. Un ejemplo es el protocolo IPv6 capaz de proveer servicios de video streaming en entornos móviles.

Una rápida visión a los protocolos MIP [13] más representativos propuestos por el IETF como son: MIPv6, HMIPv6, y PMIPv6, concluye que el protocolo MIPv6 acusa en mayor medida la

latencia del canal debido a que es necesario intercambiar un mayor número de mensajes de señalización. Por el contrario, el protocolo PMIPv6 sufre en menor medida la latencia, debido a que el MN no está involucrado el la señalización del traspaso. Las latencias MIPv6, PMIPv6, y HMIPv6 están documentadas en la RFC 2462.

Es importante señalar que el tiempo que transcurre para autentificar el MN en la red de destino añadirá cierta latencia, existen modos de disminuir la latencia ocasionada en el proceso de autentificación, para ello se aproximan las entidades de autentificación con delegación de funciones junto el gateway de acceso móvil dentro del dominio administrado. En RFC 5779 se describe el mecanismo que provee autentificación en la continuidad del servicio cuando el MN se mueve por diferentes dominios administrados.

Por lo tanto, es importante analizar la continuidad del servicio en la capa de red, hay que analizar los protocolos que intervienen en la capa de red debido a la importancia que tienen para mantener la continuidad del servicio, como ejemplo el protocolo IPv6. El protocolo IPv6 se diseña para hacer más eficiente la señalización a través de las extensiones de cabecera. Por ejemplo, existe una extensión de cabecera routing con valor 43 donde es posible almacenar la lista de nodos que sigue el paquete hasta llegar a su destino (RFC 2460). Su importancia es aún más crítica en sistemas inalámbricos y arquitecturas de acceso a red cada vez más heterogéneo. Las técnicas propuestas se estandarizan en redes celulares (TR 23.829, Septiembre 2012), el objetivo busca permitir *handovers* entre tecnología y operador. Por lo tanto, el protocolo IPv6 es crucial en tecnologías móviles, y sin una gestión eficiente de los eventos de movilidad en diferentes escenarios móviles y casos de uso no será posible proporcionar servicios de banda ancha dentro de arquitecturas de Internet móvil con razonable QoE. En este sentido es importante identificar los protocolos adecuados capaces de transformar el nivel de QoS del servicio al nivel de QoS de red.

2.3.1. Mecanismos de handover en la capa de red

Para soportar mecanismos de *handover* en la capa de red es necesario mantener la dirección IP, dirección que debe sobrevivir al cambio de red de acceso. Existen casos en los que es necesario mantener la dirección IP, por ejemplo las conexiones realizadas por el protocolo TCP, donde no es posible que los nodos participantes cambien su dirección IP durante el transcurso de una conexión. Si la dirección cambia la conexión TCP se corta si nadie responde al MN en la dirección IP que configuró el MN durante en el acceso.

En este contexto, el estándar Mobile IPv4 (MIPv4) recogido en RFC 3344 fue una de las primeras soluciones destinadas a cubrir las necesidades de movilidad de los usuarios para mantener la conectividad IP cuando el MN utiliza diferentes redes de acceso y aún cuando en el transcurso de una sesión exista un cambio de red de acceso. MIP es una solución de movilidad en capa de red que cubre tanto el handover como la gestión de la localización de los MN. MIP se fundamenta en una solución de redireccionamiento. En este sentido y como se analizará en apartados siguientes, el protocolo MIP puede enmarcarse dentro de los protocolos de movilidad IP basados en host.

El protocolo MIP evoluciona a través de múltiples complementos y extensiones, para mejorar las prestaciones del protocolo tanto en redes IPv4 como en redes IPv6. Como ejemplos más significativos en la RFC 3024 se propone una solución dereverse tunneling para MIP sobre redes IPv4 como mecanismo de optimización de ruta. En el caso de redes IPv6, en la RFC 3775 se especifica un nuevo protocolo denominado MIPv6 que ha sido especialmente diseñado para aprovechar las capacidades ofrecidas por la capa de red en IPv6.

Por otro lado, para facilitar el uso de protocolos de movilidad IP en escenarios donde coexistan redes IPv4 e IPv6, se desarrolla la extensión Dual Stack MIPv6 (DSMIPv6) a partir del protocolo MIPv6, que permite soportar tanto direcciones IPv4 como IPv6. El estándar

DSMIPv6 es uno de los soportados en la red troncal EPC en caso de acceder a los servicios mediante redes no 3GPP. El protocolo DSMIPv6 se especifica en la (RFC 5555).

Como era de prever la plataforma MIP evoluciona hacia PMIPv6 (RFC 5213) [9], más centrado que MIPv6 en los procedimientos de movilidad en entornos wireless con *handovers* transparentes para el MN, que en añadir componentes a la red, requisito exigido por el 3GPP para adoptar el protocolo en su estructura de red, debido fundamentalmente a que sus soluciones no satisfacen los requisitos de movilidad. De ese modo, el protocolo posibilita que dispositivos IP puedan moverse por redes inalámbricas desacoplando al MN de la señalización, esto significa que la red quien se encarga de señalizar el procedimiento de *handover* [10].

Es importante destacar que para dar continuidad en el servicio, también pueden plantearse otras opciones como una funcionalidad integrada para proveer servicios finales. En este sentido, y dentro del contexto de servicios soportados sobre plataformas IP se desarrolla la arquitectura *Multimedia Subsystem* (IMS) que provee accesos para el protocolo *Session Initiation Protocol* (SIP), con funciones de registro y de localización ubicadas en servidores de dicha plataforma. Las soluciones pueden constituir la base de una solución de movilidad en la capa de aplicación. Aunque la solución no soporta el protocolo MIP. Circunstancia que obliga a buscar nuevas arquitecturas capaces de soportar continuidad en la sesión, y transparentes para el MN.

2.3.2. Mecanismos para optimizar el handover en la capa de red

Los protocolos MIP analizados en el siguiente apartado (RFC 5944) como Mobile IPv6 (RFC3775), y SIP-Mobility [SIPMM] [10] proporcionan continuidad en una sesión TCP y RTP, aunque tienen el inconveniente de que no consiguen reducir la latencia de un MN cuando cambia su posición entre dominios administrados. En realidad la gestión en capas L2 y L3 implica añadir ciertos retardos. El retardo durante un procedimiento de handover es consecuencia de los procedimientos asociados a la capa de red. En primer lugar está la detección del MN, en segundo lugar la adquisición de una nueva dirección por parte del MN, y finalmente la actualización del enlace entre direcciones de red local y externa.

El procedimiento puede ocasionar un periodo de interrupción en la conectividad del MN y que sea incapaz de transmitir o recibir datos. El intervalo de interrupción se conoce como tiempo de latencia en el *handover*. Tal circunstancia obliga a diseñar técnicas para optimizar la latencia, y que el intervalo sea el menor posible, todas las técnicas están basadas en la disminución de mensajes de señalización dentro de un dominio administrado. Desde este punto de vista, surgen dos tipos de enfoque para resolver el problema de la movilidad. El primero con soluciones *micro-mobility* como HAWAII, y el segundo enfoque con soluciones *macro-mobility* como HMIPv6.

Todas las soluciones buscan reducir la latencia durante el *handover*. En paralelo a las soluciones de movilidad se diseñan esquemas de autenticación como el propuesto en [11]. Es importante diseñar eficientes mecanismos de autenticación y autorización en la nueva red de acceso para reducir la latencia que ocasiona la autorización en el nuevo dominio administrado. Así, cuando se ejecuta el cambio de red, el reestablecimiento de los servicios a través de la nueva red puede hacerse más rápido debido a que el usuario no tiene que esperar a ser autenticado en el nuevo dominio.

Otros mecanismos de mejora para tratar el problema de la movilidad IP son los mecanismos de transferencia de contextos entre redes, por ejemplo los contextos de seguridad, los mecanismos de compresión de cabeceras, y perfiles de QoS, son mecanismos que contribuyen a reestablecer de forma más rápida los servicios en la nueva red de acceso. Mediante la transferencia de contextos entre redes, la red destino puede omitir algunos procedimientos destinados a restituir dichos contextos, aunque en algunos casos los procedimientos llevan implícito un intercambio adicional de señalización entre la red y el MN, tal circunstancia añade retardos en el reestablecimiento del servicio.

Finalmente se diseñan mecanismos para proporcionar información potencial de las redes de acceso para traspasar el servicio al MN, son mecanismos que optimizan el *handover* en la capa de red. Por ejemplo, mecanismos de descubrimiento y selección de red agrupados. El documento RFC 5113 proporciona una visión detallada del alcance de los mecanismos en el contexto de la problemática de selección de red o punto de acceso. A modo de ejemplo, mediante los mecanismos de *Network Discovery* es posible proporcionar a los terminales conectados en una red de acceso la información necesaria para que puedan acceder a los servidores de preautenticación, por ejemplo direcciones IP de estos servidores RADIUS.

2.3.3. Mobile IPv6 (MIPv6)

El protocolo Mobile IPv6 [12] (RFC 3375) junto con sus extensiones permite a los host disponer de una dirección IPv6 única, dirección independiente de su punto de conexión en la red. El protocolo esta orientado a redes Internet, no se pensó desde le principio para entornos inalámbricos. El protocolo tiene un enfoque macro-mobility, es efectivo en macroceldas grandes donde los handovers son escasos. En realidad, MIPv6 no es un protocolo orientado a los handovers, sino más bien un protocolo de actualización de ruta y mantenimiento de la sesión durante el movimiento del MN. El protocolo garantiza el encaminamiento de paquetes cuando un usuario se mueve hacia una red diferente de la que estaba inicialmente. Para ello, utiliza una dirección temporal *Care-of Address* (CoA), dirección que pertenece a la red visitada por el MN, de este modo el nodo establece un túnel bidireccional con el *Home Agent* (HA), dotando de conectividad a los dos agentes mediante un túnel IPv6.

La figura 5 representa la arquitectura general del protocolo, describe el funcionamiento del protocolo en redes IPv6, donde el MN tiene una dirección estática *Home of Address* (HoA). Cuando un MN visita una red, registra una entrada en el HA. El registro incluye la CoA real tomada de la red visitada, independiente de su punto de conexión a la red Internet, y el HoA del MN, dirección inicial del MN en la red origen, de este modo el *Home Agent* siempre conoce la ubicación real del nodo móvil. Posteriormente se crea el enlace entre las dos direcciones, el enlace se actualiza de forma periódica mediante mensajes de control, mensajes de tipo *Binding Update* (BU) enviados por el MN, y reconocidos por el HA como mensajes *Binding Acknowledgement* (BA). El túnel bidireccional IPv6-to-IPv6 que se crea entre el MN y el HA mantiene viva la conexión. El Home Agent lleva a cabo la encapsulación y desencapsulación de los paquetes que pertenecen al nodo móvil, impersonalizando la presencia del MN en su Home Network. Por ultimo, el CN es el nodo al que se conecta el MN y en principio no sabe cual es su posición real, para ello el MN averigua su dirección IP a través de su HoA que es la dirección local del CN.

Hay que tener en cuenta que la operación anterior establece una ruta generalmente sub-óptima que contiene la HA dentro del enlace MN-CN. La ruta se usa para la comunicación entre el HA, y MN, los paquetes enviados por el CN al MN son encaminados por el HA. Este fenómeno denominado encaminamiento triangular introduce latencias adicionales y overhead de red, aunque puede ser eliminado registrando directamente el MN en el CN, mediante un par de mensajes Binding Update, mensajes de reconocimiento. Por supuesto, esto requiere que el CN tenga capacidad para entender el protocolo MIPv6 y también capacidad para utilizar algunos mecanismos de seguridad adicionales, como identificar del CN con alguna certeza razonable de que el MN es efectivamente direccionable desde su CoA así como la HoA, el procedimiento de autentificación básica, conocido como *Return Routability Procedure* (RRP) (HoTI-CoTI-HoT-COT) debe ser ejecutado antes de la secuencia BU/BA. (Ver figura 5).

Mientras Mobile IPv6 con sus extensiones de seguridad es una solución viable para proporcionar una conectividad para los terminales en movimiento, hoy en día sin embargo, todavía se utiliza IPv4 mayoritariamente en Internet, de este modo la información entregada no sería eficiente sin un mecanismo de transición IPv4-IPv6. Mecanismo denominado *Dual-Stack Mobile IPv6* (DSMIPv6) (Soliman H., 2009), esta es una de las técnicas que extienden la

funcionalidad de MIPv6 a la presencia de IPv4 en las redes de acceso, sin embargo, debido a su complejidad no se utiliza ampliamente.

Lo que se pretende conseguir en ambos casos es ofrecer al usuario una red funcional, capaz de mantener las conexiones activas todo el tiempo, independientemente de la posición del nodo, y el protocolo utilizado dentro de un dominio macro-mobility.

Por ejemplo, un escenario típico de uso ocurre durante el establecimiento de una llamada VoIP. Cuando un CN quiere contactar con un MN para establecer un servicio VoIP, lo que hace es intentar conectar con el MN a través de su HoA, que es la dirección fija conocida por el CN. Los paquetes enviados a la red del operador dirigidos al HoA del MN, son interceptados por el HA, encapsulados en un paquete MIPv6 y redirigidos hacia la nueva dirección del CoA que corresponde con la del nodo móvil en la red visitada.

El nodo móvil contesta al CN encapsulando los paquetes de datos en un paquete MIPv6, paquete que envía al HA el cual extrae el paquete original del paquete MIPv6 recibido, enviándoselo al CN.

En el caso de que el CN tenga configurado una dirección IPv6, el MN pude adicionalmente contactar con el CN para informarle de su dirección IPv6, de ese modo el CN envía los paquetes directamente al MN, el procedimiento como se ha comentado en párrafos anteriores se denomina "proceso de optimización de ruta", siendo una mejora en la ruta seguida por los paquetes puesto que no tiene que pasar por el HA, que introduce retardos innecesarios. En el caso de que el CN no soporte IPv6, el invento no es posible, no es posible que el CN y el MN puedan utilizar el proceso de optimización de ruta. Finalmente, en el caso de que el MN cambie de red, este obtendrá una nueva CoA que deberá registrar en el HA con el fin de que sea alcanzable por el CN.

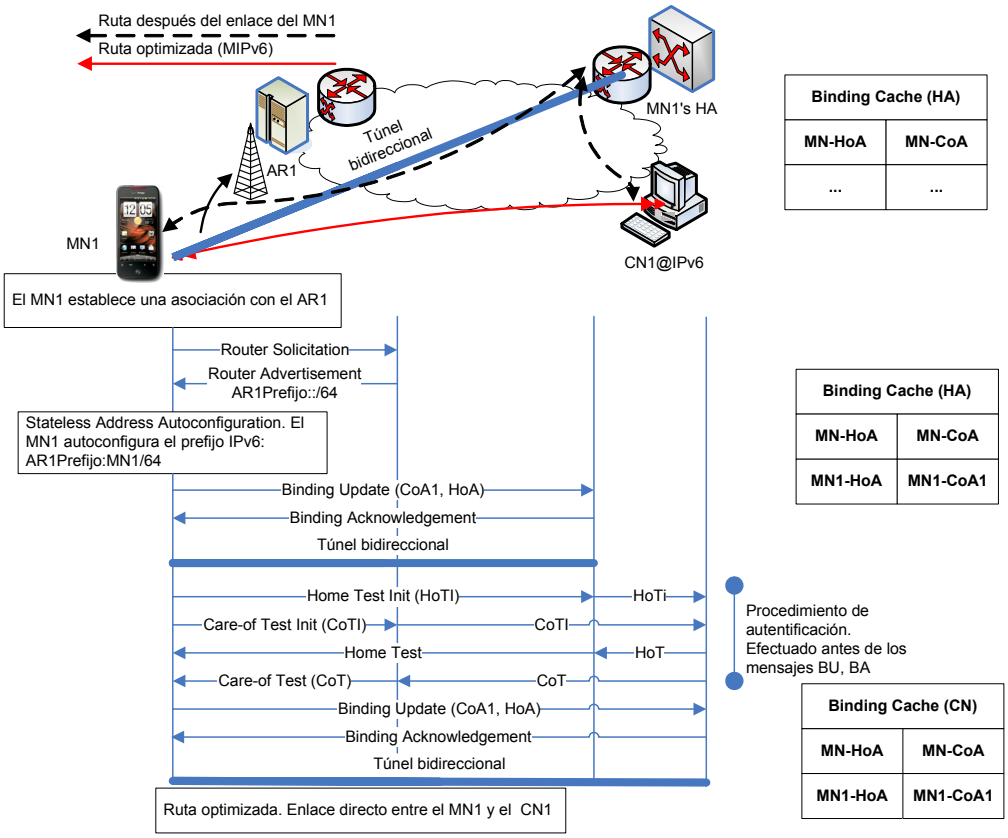

Figura 5. Arquitectura básica de Mobile IPv6

MIPv6 es más eficiente que MIPv4 para transmitir un servicio como IPTV, dado que el retardo es menor y la perdida de paquetes es menor durante la transmisión. La perdida en MIPv4 es del 7%, mientras que en MIPV6 al no existir el Foreign Agent (FA) y la existencia de

procedimientos de comunicación directa la perdida es solo del 1%. Hay que tener en cuenta que los componentes funcionales de ambos protocolos son diferentes, por ejemplo en el caso de MIPv4 son: el MN, HA y FA. En este caso el HA y el FA emiten anuncios a la red para que el MN pueda saber en que red se encuentra, el en caso del protocolo MIPv6 las entidades funcionales son el HA, CN y MN.

Para cumplir los requisitos de seguridad RRP un nodo móvil cualquiera recibe un mensaje relacionado con su HoA, y su CoA. El MN debe comprobar que realmente corresponden al nodo. Tal como se ha descrito en párrafos anteriores, el HA y el CN son capaces de procesar mensajes de tipo Binding Update (BU). Los mensajes BU desde el MN al HA están protegidos mediante IPSec, utilizan un mismo dominio administrado, circunstancia que permite incluir los campos de cabecera Authenticator Header (AH), y un Encapsulating Security Payload (ESP) mediante un mecanismo de seguridad preestablecida, esto significa que solamente el HA procesa los mensajes BU, con lo cual es fácil asumir que debe existir una seguridad preestablecida entre ambos mediante certificados, clave simétrica, etc.

Los mensajes BU desde el MN al CN deben seguir un patrón diferente, no es posible utilizar el método anterior, esto es debido por que a priori el MN no sabe el CN destino. Asumir en ese caso una relación de confianza con todos los CNs es inadmisible. El método se basa en verificar que el nodo que es alcanzable a través de la HoA es el mismo nodo que es alcanzable a través de la CoA. Para ello cuando un MN desea enviar un mensaje de tipo BU a un CN, debe obtener un modo de validar dicho mensaje. El procedimiento se realiza del siguiente modo, el MN solicita al CN que le envié una clave al HA. De otro modo, el MN solicita al mismo CN que le envié una clave a su CoA. Una vez que el MN dispone de las dos claves realiza una función de hash sobre ambas, el MN añade información adicional al mensaje, y genera un mensaje de autorización BU que envía al HA y al CN.

Cuando el CN recibe el mensaje de autorización BU, verifica la información adicional antes de procesarlo. A partir de esa información, el CN puede verificar que el MN es accesible desde el HA, y su CoA.

2.3.4. Hierarchical Mobile IPv6 (HMIPv6)

HMIPv6 [13] es una extensión de la estructura del protocolo MIPv6, mejora MIPv6 reduciendo la latencia en el handover, permite estratificar jerarquías al proceso asociado de movilidad IP. De ese modo, se disminuyen latencias durante el traspaso, debido a que la actualización del movimiento en un área jerárquica se realiza en menor tiempo. Los movimientos del MN dentro de un mismo dominio administrado son transparentes para el CN y el HA. El protocolo opera mediante un enfoque basado en *micro-mobility* dentro del dominio administrado, y *macro-mobility* dentro del dominio donde se ubica el HA. HMIPv6 reduce la señalización respecto MIPv6, es capaz de reducir la señalización en el interfaz radio. Escala mayor número de MNs que MIPv6, finalmente proporciona optimización de ruta.

En el diseño del protocolo HMIPv6 se introdujo un nuevo nodo de red, *Mobility Anchor Point* (MAP), que tiene la misma funcionalidad que el HA, es capaz de guardar vinculos entre dos direcciones IPv6, la dirección local y la dirección remota. En HMIPv6 se utilizan dos tipos diferentes de direcciones, *Regional Care-of Address* (RCoA), y *On-link Care-of Address* (LCoA), esta última tiene la misma funcionalidad que la CoA en MIPv6, su nombre LCoA solamente persigue ser un nombre distinto al de RCoA. El RCoA es una dirección de la subred de la MAP.

La llegada de un MN en un dominio HMIPv6 implica que el MN autogenera una dirección IPv6 a partir de los anuncios de ruta del router de acceso (AR), es la dirección LCoA. Los anuncios de ruta contienen información de los MAPs existentes en el dominio. En el caso de que solamente exista un MAP el MN puede decidir si utiliza HMIPv6 ó simplemente MIPv6 con la LCoA generada. Si decide utilizar la HMIPv6, el MN solicitará una RCoA al MAP.

Después enviará un mensaje de señalización local BU, en el que incluye las dos direcciones adquiridas, la LCoA y la RCoA. El MAP procesará el mensaje BU, almacenará las dos direcciones en su Binding Cache, en ese momento actúa como si fuera un HA para la dirección RCoA, esto es intercepta todos los paquetes enviados a esa dirección, y los reenvía a la posición real del MN. El MAP envía un mensaje de tipo Binding Acknowledgement de vuelta al MN. Se crea un túnel bidireccional entre el MAP y el MN. Finalmente el MN envía un mensaje BU a su HA con la dirección RCoA en el campo CoA (RCoA=CoA).

Cuando un MN se mueve entre nodos dentro de un domino MAP, solamente cambia de LCoA que le proporciona el AR y registra de nuevo en el MAP, por el contrario si se mueve entre dos MAPs se genera tanto la LCoA como RCoA, y actualiza la entrada en el HA. En un dominio gestionado por el MAP los handovers se manejan localmente, sin necesidad de enviar mensajes de señalización a nodos HA que quizás estuvieran lejos el domino MAP. De ese modo, el movimiento del MN es transparente para todos los nodos que están fuera del dominio administrado. La figura 6 ilustra el procedimiento cuando llega un MN a un dominio HMIPv6, y empieza una sesión con el HA. El MN se mueve entre dos AR dentro de un dominio MAP. El MN envía un mensaje BU al MAP la dirección LCoA y su RCoA.

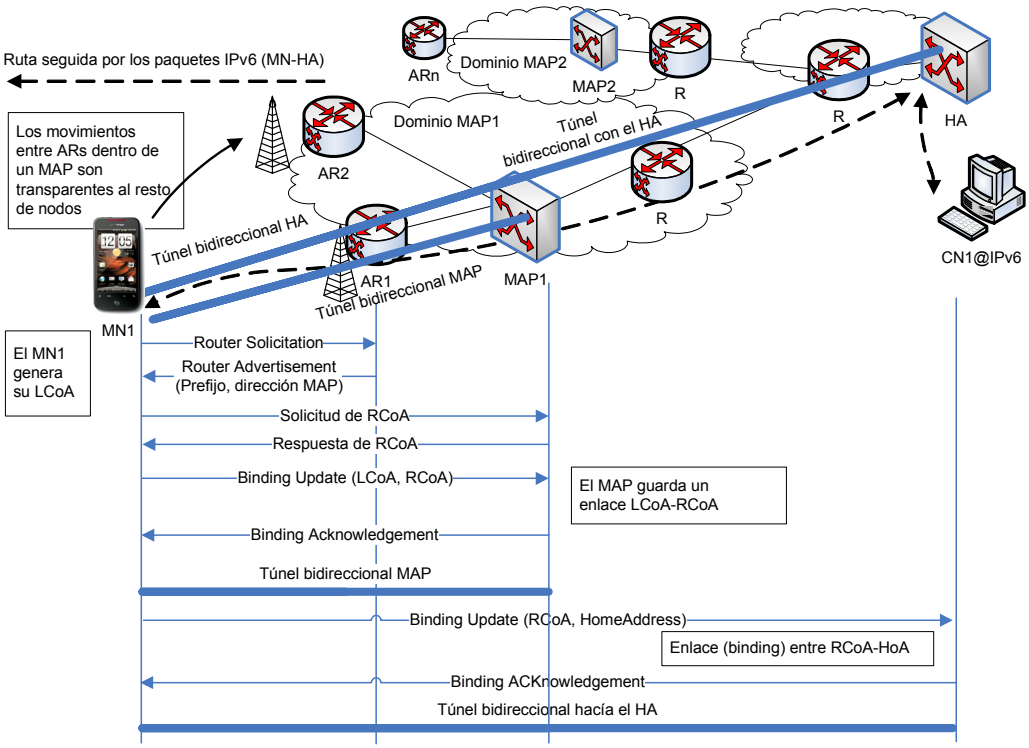

Figura 6. Señalización de un túnel MN-HA en HMIPv6

2.3.5. Mobile IPv6 Fast handovers (FMIPv6)

El protocolo FMIPv6 (RFC 5568), es una extensión del protocolo MIPv6 e independiente de protocolos de capa 2. FMIPv6 realiza handovers rápidos, consigue disminuir los paquetes perdidos durante el movimiento del MN en los diferentes ARs que encuentra el MN en las redes de acceso. FMIPv6 es un mecanismo fabuloso para disminuir el overhead, y la latencia en el handover. Llegando a 20ms de latencia máxima algunos casos. Los mecanismos QoS se gestionan a través de esquemas de modelos de servicios, básicamente a través de DiffServ.

El protocolo predice el próximo AR que se encontrará el MN durante su movimiento, de esta forma es posible obtener una nueva IPv6 antes de que el MN establezca una sesión en el nuevo AR (NAR). El segundo objetivo del protocolo consiste en utilizar el AR actual (PAR) para reenviar los paquetes dirigidos al MN que se encuentra ubicado en el NAR, se reduce de ese modo el número de paquetes perdidos durante el handover.

En FMIPv6 se define un nuevo mensaje, *Router Solicitation for Proxy Advertisement* (RtSolPr), que es enviado por el MN a su AR (PAR) para obtener información sobre los ARs adyacentes. El PAR responde con un mensaje *Proxy Router Advertisement* (PrRtAdv). El MN elige el NAR más apropiado de la lista que le proporciona el PAR, y genera la una nueva CoA (NCoA), de acuerdo al prefijo utilizado en la red del NAR, previamente mediante anuncios Neighbor Discovery comprueba que la dirección no este asignada. Finalmente, el MN envía un mensaje de sincronización Fast BU (FBU). El protocolo propone dos modos de operación, esto es el handover predictivo y el handover reactivo.

El método predictivo consiste en que el MN envíe un mensaje FBU desde su red de acceso anterior PAR, esto es la red antes de realizar el handover, y esperar allí a que le envíen un mensaje de tipo Fast BAck (FBack). El PAR, envía un mensaje handover Initiate (HI) al nuevo AR. El NAR reconoce el handover al PAR con un mensaje de tipo Handover Ack (hack). El PAR genera y envía un mensaje de tipo FBACK hacia el MN y HAR después de recibir el mensaje HAck. En paralelo se inicia el reenvío de paquetes del PAR dirigidos al MN ubicado en la nueva red. Después que el MN llega a la nueva red envía un mensaje Unsolicited Neighbor Advertisement (UNA) hacia su NAR, y a partir de entonces el MN puede recibir inmediatamente sus paquetes desde el CN.

En el método reactivo no es necesario enviar un mensaje FBU desde la red anterior, inmediatamente después de recibir el PrRtAdv, el MN puede unirse en el NAR enviando un mensaje de tipo UNA al NAR. En el siguiente instante de tiempo, el MN envía un mensaje de tipo FBU al PAR quien inicializa el handover con el método anterior, esto es, el NAR reconoce al MN mediante un HI, y lo confirma con un HAck. En ese momento se lleva a cabo el traspaso. En ambos casos durante el traspaso, los paquetes se reenvían desde PAR a la MN a través de NAR. Por razones de rendimiento los protocolos HMIPv6, y FMIPv6 habitualmente se utilizan juntos [14].

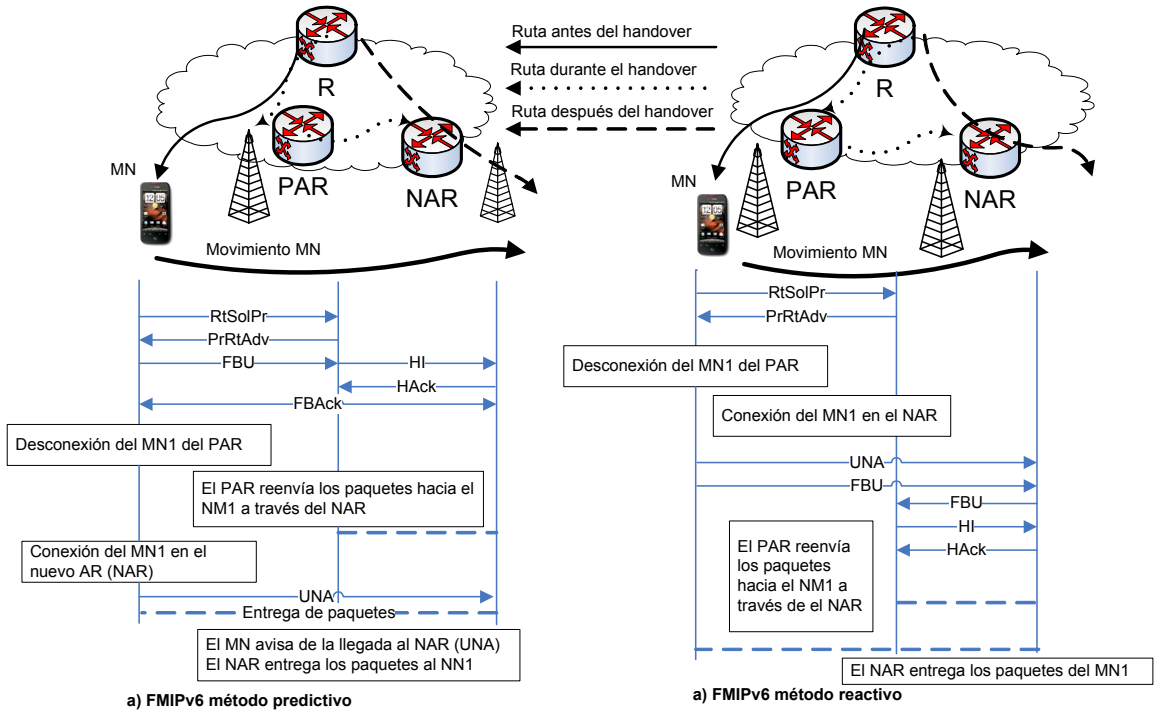

Figura 7. FMIPv6, arquitectura y handovers entre nodos

En realidad se utiliza HMIPv6, y FMIPv6 de forma conjunta [15]. La idea es convertir al MAP en un nodo convergente o divergente (depende de la ruta) dentro de un dominio MAP, siendo el nodo que gestiona la micro-movilidad. Dentro del dominio MAP se utiliza FMIPv6, durante los handovers se utiliza un túnel para reenviar los paquetes desde el antiguo router de acceso al nuevo router de acceso. El mecanismo ayuda a reducir la pérdida de paquetes, aunque el

mecanismo no es suficiente para aprovisionar QoS, debido fundamentalmente a que en las redes inalámbricas son altamente dinámicas, esto es los usuarios están en constante movimiento. Circunstancia que requiere de señalización para aprovisionar recursos para conseguir niveles adecuados de QoS, como vemos a continuación.

2.3.6. Solución MEMO MCoA (Multiple Care-of Address)

Los protocolos de movilidad necesitan una excesiva señalización. También tienen el inconveniente de que todas las conexiones deben realizarse a través de un único interfaz de red, ya sea 3G, 4G o WLAN.

De ese modo, para que un MN pueda conectarse a varias redes de acceso de forma simultánea, utilizando interfaces diferentes es necesario introducir nuevas extensiones al protocolo MIPv6 (*Multiple Care-of-Address* (MCoA)) (RFC 5648). Extensiones MIPv6 que mejoran la latencia de red, el overhead, y manejan de forma más humana las políticas de enrutamiento.

La figura 8 ilustra la arquitectura de MEMO MCoA. El MR tiene dos interfaces externos, enlazados con redes de acceso diferente a través de una única CoA. Utilizando la CoA se crea un túnel MIPv6 hacia el HA, el HA establece el vínculo en el extremo final del túnel para asociar la HoA-CoA. Mientras que con NEMO BS para crear el túnel IPv6-to-IPv6 solamente se utiliza una CoA y una HoA, no sucede lo mismo para MEMO MCoA, esto es debido a que en el final del túnel el HA tiene que discriminar e identificar los diferentes túneles. Le sucede lo mismo al MR debe saber como encaminar los datos hacia un túnel individual a través del interfaz físico.

Tal circunstancia se debe a que el MR no tiene información suficiente de capa de red para decidir sobre que túnel encaminar los datos. Una solución consiste en introducir un nuevo identificador llamado Binding Identifier (BID). El identificador identifica el interfaz sobre el que se construye el túnel. El BID se envía al HA a través de mensajes de señalización BU, de este modo el HA es capaz de identificar los flujos en el extremo final del túnel.

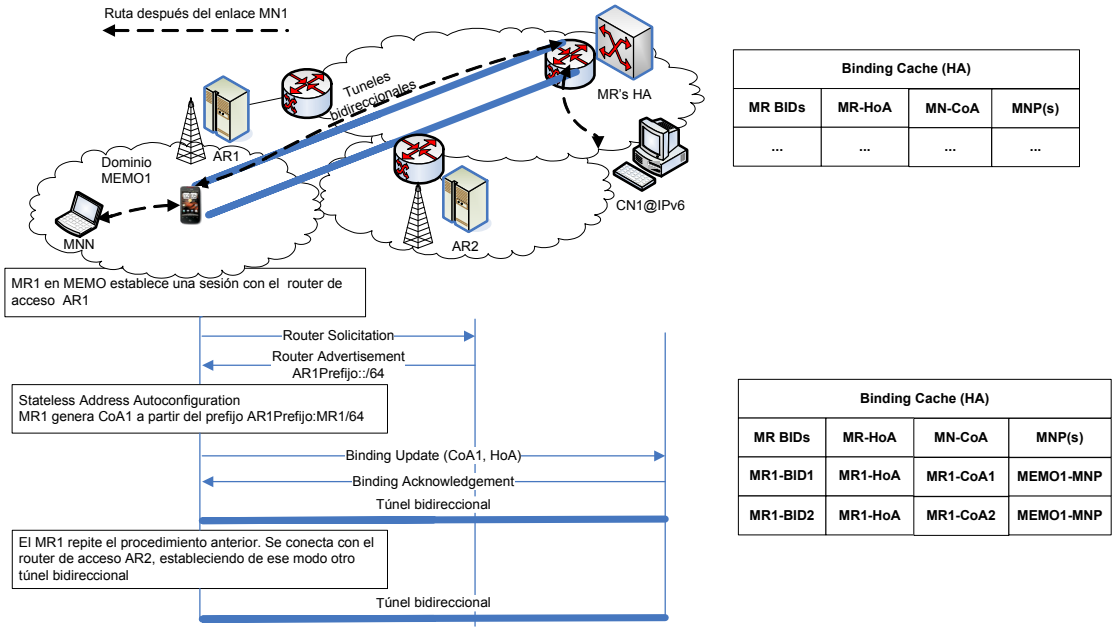

Figura 8. MENO multi-interfaz

Al MR le sucede lo mismo, identifica a través de identificadores BID el el túnel. Es capaz de encaminar paquetes evitando tráfico asimétrico. El objetivo es evitar tráfico asimétrico, que los paquetes que pertenezcan a diferentes flujos sean encaminados por túneles diferentes.

El MR envía al HA el identificador de túnel BID. El HA, almacena el identificador y es capaz de encaminar paquetes de un determinado flujo por el interfaz correspondiente (RFC 6088).

La solución anterior introduce el concepto de multihoming, solución válida para encaminar diferentes flujos móviles IP, y mejorar la experiencia de usuario durante el movimiento del terminal. También es una solución válida para que un MN pueda establecer una sesión con diferentes interfaces utilizando tecnologías diferentes, de este modo la entrega de datos sobre todo multicast es más fiable y robusta.

2.3.7. Proxy Mobile IPv6 (PMIPv6)

Mobile IPv6 proporciona movilidad IP en servicios seamless en entornos reales, a pesar de que la latencia introducida que es elevada. El protocolo en entornos inalámbricos no es demasiado eficiente, básicamente debido al retraso en términos de latencia, y perdida de paquetes que causa el procedimiento de traspaso. Retraso inaceptable en servicios en tiempo real. Esto es debido a que MIPv6 fue diseñado como solución basada en host, donde la señalización y los túneles son inicializados por el MN, y en algunas situaciones el MN se encuentra en una red de acceso. Por otro lado, el protocolo requiere que se modifique para su implementación el software del dispositivo móvil o fijo, situación que introduce importantes problemas, en algunos casos simplemente no es posible. De este modo, el despliegue de MIPv6 en los nuevos dispositivos es muy lento. Hay que tener en cuenta que la implantación del protocolo MIPv6 en el kernel del dispositivo presenta algunos problemas de seguridad.

Para evitar, y dar solución a estos problemas el IETF crea el grupo de trabajo *Network-Based Localized Mobility Management* (NetLMM), encargado de buscar nuevas soluciones de movilidad IP. El resultado son soluciones basadas en red, los handovers se gestionan automáticamente por la red sin necesidad de involucrar al MN. Tal circunstancia permite ahorrar señalización. La idea es dejar que el MN mantenga su dirección IPv6 durante los movimientos a través de los diferentes routers de acceso para conseguir de ese modo que la movilidad sea transparente en la capa IP. Otro beneficio que se consigue es hacer agnóstica el traspaso a la tecnología.

El NetLMM propone como solución *Proxy Mobile IPv6* (PMIPv6) (RFC 5213), el Proxy gestiona los handovers producidos por el MN, y consigue que sean transparentes para estos, mejora también la seguridad durante el proceso de autentificación en el handover, y optimiza la QoS durante el proceso de automatización de rutas.

El protocolo, tiene ciertos retrasos inevitables, debido a la necesidad de autentificarse en el nuevo router de acceso, o el tiempo que tarde la autorización en ese nuevo router. En cualquier caso, los retrasos son mucho menores que en MIPv6. Esto es debido a que el MN se autentifica solamente en el gateway móvil del dominio de acceso PMIPv6.

PMIPv6 es transparente al dispositivo e independiente de la tecnología, tal situación consigue hacer agnóstico el protocolo a tecnologías como WLAN, WiMAX, 3G UMTS, LTE, LTE-A o cualquier otra tecnología de futuro. Provee fast handovers dentro de un área determinada, lo que significa que PMIPv6 tiene un dominio bien definido, el Local Mobile Domain (LMD), donde el intercambio de mensajes de señalización es bastante rápido. PMIPv6 garantiza una misma dirección IP al MN durante el movimiento, de ese modo proporciona continuidad durante la sesión dentro del dominio administrado de red: “dominio PMIP”. Tal circunstancia favorece a los protocolos de capas superiores, debido a que no deben construir nuevas sesiones después de un traspaso porque la dirección IP y los puertos del protocolo de transporte siguen siendo los mismos.

El protocolo PMIPv6 define nuevas entidades lógicas. El nodo *Local Mobility Anchor* (LMA) y el nodo *Mobile Access Gateway* (MAG). El LMA actúa como un home agent en MIPv6, mantiene el conjunto de rutas de todos MN dentro del LMD y todo el tráfico dirigido hacia los MNs tiene que pasar por este nodo. El LMA almacena en su Binding Cache (BC) una entrada *Home Network Prefix* (HNP) por cada MN junto al identificador del MN. Tablas que necesitan actualizarse periódicamente.

El MAG es el router de primer salto ó router de acceso de los MNs. Una vez que el MN entra dentro de un dominio administrado LMD, la MAG reconoce al MN por medio de anuncios de reconocimiento de ruta, o *tiggers* de capa 2. Entonces, el MAG envía al LMA un mensaje de tipo *Proxy Binding Update* (PBU) que contiene el identificador del MN y el identificador de la MAG. El LMA responde con un *Proxy Binding Acknowledgement* (PBA) que incluye el *Home Network Prefix* (HNP) del MN para que el MN configure su interfaz. El LMA crea, entonces un *Binding Caché Entry* (BCE), esto es una entrada en la caché del LMA. La entrada en la BCE contiene los campos, MN-ID, MAG Proxy-CoA y MN prefix. Finalmente, se establece un túnel bidireccional entre el LMA-MAG para enviar el tráfico hacia los MN asociados a ese MAG. Solamente existe un tunel IPv6 entre el MAG y LMA, el tráfico hacía los diferentes MN se discrimina por las KEYs GRE.

De ese modo el MAG posee toda la información para emular el comportamiento de un MN. Envía periódicamente mensajes de descubrimiento de ruta (*Router Advertisement* (RA)) para que los MN puedan anclarse en los diferentes routers de acceso dentro del dominio LMD. Visto así, las tres principales entidades tienen establecida la siguiente configuración, el MN tiene un identificador de red válido para su home network en el punto de anclaje actual dentro de la red y el LMA-MAG contiene información suficiente para gestionar el tráfico enviado hacia o para el MN utilizando una dirección de su home network. El LMA – MAG envían el tráfico a través de un túnel bidireccional túnel que compartido entre diferentes MN.

El MAG actúa como router por defecto dentro del dominio administrado, cualquier paquete que el MN envíe al CN es interceptado por el MAG, y enviado al LMA a través de un túnel bidireccional. El LMA en el otro extremo del túnel, quita la cabecera del túnel, y encamina el paquete hacia el destino, el CN.

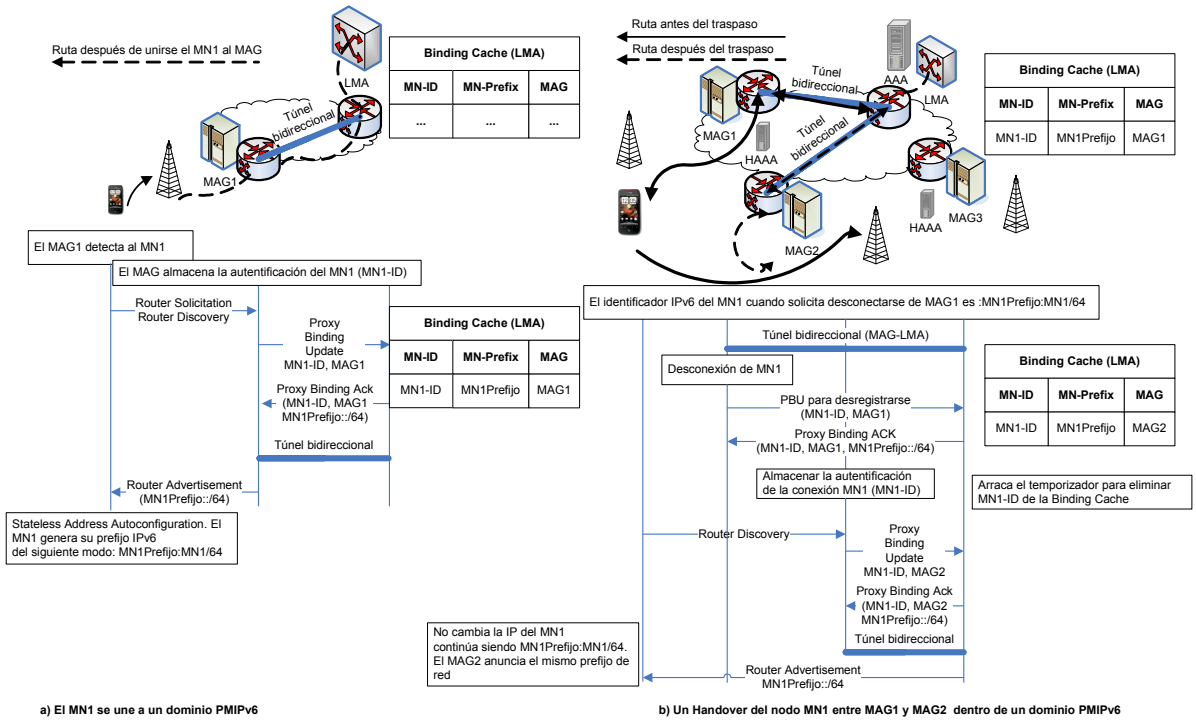

Figura 9. Arquitectura PMIPv6

El proceso de autentificación en PMIP es el siguiente, una vez que llega un mensaje de solicitud de ruta por parte del MN el nuevo MAG2 responde al MN con un mensaje que contiene un nonce. El MN envía al nuevo MAG2 el nonce encriptado con *Security Association* (SA) que obtiene del servidor de autentificación del anterior dominio PMIP, esto es MAG1 (AAAH). El MAG2 entrega el nonce encriptado al LMA/AAA. Los servidores de autentificación AAA, y AAAH se intercambian la información de modo seguro, a través de

procedimientos de infraestructura. Finalmente, el AAAH compara la información que recibe del AAA con la entregada por el MN, si son iguales el AAAH autentica al MN en el MAG2.

La figura 9 ilustra dos escenarios posibles en una arquitectura PMIPv6, representa un único LMD desde el punto de vista de los MNs moviéndose entre diferentes MAGs. La figura 9a, representa la señalización cuando un MN llega a un LMD, y este se une a la MAG más cercano. La figura 9b, representa la señalización durante el handover en el interior de un dominio PMIPv6. La solución es fabulosa y prometedora tal es así que el 3GPP la aprobó para ser candidata en sus arquitecturas de futuro.

2.3.8. Interfaces 3GPP basados en PMIPv6

La implementación basada en GTP [8] proporciona las funciones de creación, eliminación, modificación y cambio del servicio portador en el plano de usuario entre el S-GW y P-GW. Una implementación basada en PMIPv6, no contempla ninguna gestión de servicios portadores entre el P-GW y S-GW. De ese modo, en la implementación PMIPv6 el servicio portador EPS de LTE se extiende desde el equipo de usuario hasta el S-GW y no hasta la pasarela P-GW, como sería el caso de utilizar GTP en la interfaz S5/S8. Ello es debido a que el protocolo PMIPv6 se diseño para ofrecer exclusivamente un servicio de movilidad entre el S-GW y el P-GW y no dispone de los mecanismos necesarios para señalizar parámetros de QoS.

Por tanto, las interfaces S5 y S8 cuando utilizan la variante basada en PMIPv6 [9], ofrecen únicamente un servicio de conectividad entre las pasarelas, de modo que todos los flujos de tráfico se transportan en un mismo túnel sin distinción de servicios portadores.

La implementación basada en PMIP añade al S-GW la funcionalidad de agente relay para DHCPv4, y DHCPv6. La implementación lleva a cabo los mensajes de Router Solicitation, y Router Advertisement definidos en la RFC 4861.

El protocolo PMIPv6 fue adoptado por parte del 3GPP en las interfaces S5/S8 entre las pasarelas S-GW y P-GW como alternativa de uso del protocolo GTP. El protocolo GTP fue el protocolo inicialmente especificado por el 3GPP. Ambos protocolos, GTP y PMIPv6 resuelven la movilidad de modo transparente para el MN (TS 23.402)

La dirección IP de envío de paquetes es la dirección del router que lleva a cabo la funcionalidad MAG. Para el interfaz S5/S8, todos los paquetes IP que llegan a la pasarela P-GW (LMA) desde la red externa son paquetes que contienen como dirección destino la dirección IP asignada al MN. Los paquetes son encapsulados y posteriormenente enviados mediante un túnel IPv6 a la pasarela S-GW que contiene el MAG embebido y viceversa. La transferencia de paquetes IP entre las pasarelas S-GW y los dispositivos de usuario (MN) no corresponde a protocolos de movilidad, se lleva a cabo mediante procedimientos en el interfaz S1 y mecanismos radio mediante una arquitectura de servicios portadores.

En el plano de control el protocolo intercambia mensajes de señalización entre el MAG y el LMA. A través de dichos mensajes de señalización se gestionan y establecen las asociaciones de direcciones y los túneles necesarios para el envío de datos. En el plano de usuario el protocolo señaliza el establecimiento de un túnel que permite enviar de forma transparente los paquetes IP de los usuarios, con las direcciones origen y destino pertenecientes al espacio de direcciones de la red local y externa para el usuario los interfaces S5/S8, de ese modo el MAG y LMA son transparentes. Las direcciones pueden estar ubicadas en un espacio de direcciones diferente en la red externa.

El trasporte se realiza encapsulando los paquetes IP con el protocolo GRE (*Generic Routing Encapsulation*, (RFC 2784)), protocolo capaz de identificar flujos. El protocolo GRE asocia cada paquete con la conexión PDN a la que pertenece. Existe una diferencia respecto a GTP, los túneles GRE utilizados por PMIPv6 no distinguen el tráfico que pertenece a diferentes servicios portadores dentro del EPS y que establecen un contexto en una misma conexión

PDN. Esto significa que los identificadores de túnel GRE hacen referencia a una conexión PDN mientras que los identificadores de túnel GTP-U hacen referencia a un servicio portador EPS. Esta diferencia hace que los servicios portadores EPS, según se comento en el apartado anterior se extiendan entre la pasarela S-GW y el equipo de usuario, cuando se utiliza PMIPv6 en la interfaz S5/S8. Por el contrario, mediante el uso de GTP en S5/S8 los servicios portadores EPS se extienden desde la pasarela P-GW hasta el terminal de usuario (MN).

Finalmente, en el interfaz S5/S8 el protocolo PMIPv6 es el protocolo especificado para soportar la movilidad entre LTE y redes no 3GPP. Las interfaces S2a, S2b están basadas en PMIPv6, interfaces que conectan accesos 3GPP con WLAN. La interfaz S2c está basada en DSMIPv6, interfaz que conecta directamente el MN con el P-GW. A partir de la Release 10 este interfaz soporta flujos móviles IP.

3. Optimización de QoS en el proceso de ruta: propuestas 3GPP e IETF

En este apartado analizamos la evolución de los mecanismos de optimización de QoS en el proceso de optimización de ruta. Son modelos matemáticos que modelan o estiman la percepción humana, en este sentido la ITU define QoS en la recomendación E800 como: “los requerimientos que debe cumplir un medio para trasportar un flujo de información”. Así pues, la disponibilidad del servicio está influenciada por el acceso de red y su arquitectura. La accesibilidad está caracterizada por la capacidad de identificar al MN y por los protocolos encargados de señalizar una ruta. Protocolos diseñados para mejorar el rendimiento del servicio en términos de throuthput, latencia y jitter. Por otro lado, la gestión de la QoS está caracterizada a través de variables como el retardo, conseguir la menor perdida de paquetes posible, maximizar la utilización de la red, entregar ancho de banda dedicado, el ancho de banda que demanda el servicio y la posibilidad de priorizar el tráfico, especialmente aquel tráfico sensible al retardo.

Una solución de movilidad IP con gestión de QoS en redes inalámbricas debe integrar el protocolo MIP. La integración obliga a buscar protocolos capaces de reservar recursos para aprovisionar diferente parámetros de QoS según las características del servicio, como servicios *best-effort* o los servicios en tiempo real.

Los protocolos de aprovisionamiento permiten escalar el servicio a través de los diferentes dominios administrados. Aunque los protocolos deben adaptarse a medios inalámbricos, y también deben adaptarse al cambio dinámico que sufre su identificador de red durante el movimiento, en este contexto la gestión de la movilidad y la QoS son independientes la una de la otra.

En este sentido se desarrollan y proponen nuevos protocolos como *Mobile Resource Reservation Protocol* (MRSVP), el protocolo actúa en modo pasivo solicitando recursos de red para poder establecer una sesión, para ello el emisor y el receptor envían mensajes de tipo MSpec en la que ambos negocian los recursos que van a aprovisionar, así hasta conseguir completar la ruta que forman los nodos de red. El mecanismo es poco eficiente en términos de optimización recursos de red, esto es debido a la posibilidad que quizás se reserven recursos que no son necesarios para aprovisionar la ruta del servicio, y muy posiblemente el MN nunca se llegue a utilizar.

Por otro lado, el movimiento del MN es determinista aunque parezca lo contrario, aunque en servicios en tiempo real es difícil de utilizar tal aproximación, tal circunstancia afecta al dimensionado de arquitecturas de red. De ese modo, aparecen nuevos esquemas para aprovisionar recursos de modo progresivo donde en MN que actúa como proxy enviando mensajes a sus vecinos para informar de un posible traspaso, y de ese modo los nodos vecinos son capaces de reservar los recursos necesarios para aprovisionar el servicio.

Para mejorar los mecanismos anteriores en entornos Wireless se desarrollan esquemas de aprovisionamiento inalámbrico *Wireless Lightweight Reservation Protocol* (WLRP) con el único objetivo de aumentar la probabilidad de que un MN visite un área determinada, y aprovisionar recursos según una probabilidad. Los mecanismos están basados en un perfil de movilidad y un perfil de servicio. De ese modo una entidad central monitoriza el perfil del servicio y envía de modo pasivo/activo mensajes a los nodos para que reserven recursos, el mecanismo es especialmente útil entornos de movilidad.

Es un mecanismo excelente para gestionar y proveer recursos para servicios E2E entre diferentes dominios administrados. El mecanismo está basado en métricas *One-way Delay Metric* (RFC 2679), técnicas útiles para medir el retraso E2E a partir de una marca de tiempo, *timestamp*, marca que envía el origen al destino. Es necesario que existan mecanismos de sincronización entre ambos agentes.

En este sentido aparece *In-Network Automatic Management Mechanism* (I-NAME), mecanismos que utilizan técnicas data-minig para fortalecer y mejorar la QoS a partir de preferencias de usuario y la disponibilidad de recursos de los red. Los mecanismos surgen debido a la compleja gestión de la movilidad influenciada por factores diversos, heterogéneos y complejos. Factores no relacionados entre si como el ancho de banda del servicio, las características de la red HetNet, la complejidad de la red y los protocolos de movilidad que se utilizan, ó las características del servicio.

Por lo tanto, para soportar QoS estricta en redes heterogéneas los protocolos deben ser capaces de solicitar, e interpretar información de los recursos disponibles en la red y de las necesidades del servicio. De ese modo a partir de un perfil de usuario y un perfil de red, la infraestructura debe ser capaz de aprovisionar recursos para el servicio. Tal circunstancia se consigue con la introducción de nuevas entidades lógicas que almacenen y proporcionan información en tiempo real a través de técnicas data-mining. En definitiva, se busca aprovisionar los recursos que el usuario espera de la red.

3.1. Mecanismos 3GPP para estructurar el tráfico

En este apartado se analiza la gestión de flujos contenedores de tráfico IP dentro de un mismo contexto PDN, se describen mecanismos capaces de mantener la continuidad en la sesión IP. Se analizan procedimientos que proporcionan movilidad IP al MN reduciendo de forma significativa el RTT. Mecanismos que han surgido en los últimos años debido a que son capaces de proporcionar una gestión eficiente del tráfico a un menor coste para el operador, descargando el tráfico por la pasarela de acceso, descongestionando la macrocelda.

Las arquitecturas de acceso Wireless de nueva generación se caracterizan por coexistir diferentes tecnologías en las redes de acceso dentro de un backbone común. El diseño asegura la posibilidad de añadir mecanismos capaces de estructurar el tráfico dentro del EPC. En este sentido, EPS LTE separa las diferentes redes de acceso del core. Dicha separación permite desarrollar mecanismos para la gestión del tráfico, uno de ellos es el control de admisión, la autorización, o hacer que el tráfico fluya solamente por las redes de acceso. El objetivo último siempre es liberar parte del tráfico de la macrocelda.

El tráfico se transporta mediante flujos, de ese modo el operador asigna reglas para cada tipo de tráfico a partir de un perfil de servicio, tal circunstancia permite descargar el tráfico de la macrocelda entre las diferentes redes de acceso, por ejemplo en caso de congestión el tráfico puede moverse por redes de acceso non-3GPP, el 3GPP denomina al mecanismo *traffic offload*. Esto significa que existe la posibilidad de mover parte del tráfico, según preferencias de usuario hacia redes WLAN, manteniendo la continuidad en la sesión con la mínima interacción posible del usuario, el tráfico en la red de acceso deberá viajar encriptado, y conservar la dirección del MN.

También es importante caracterizar el tráfico en términos QoS. De ese modo, la continuidad en la sesión en el servicio sigue siendo una parte importante en el diseño de las arquitecturas propuestas por el 3GPP.

En ese sentido, el 3GPP [17], estandariza nuevos mecanismos para estructurar el tráfico. Soluciones *como Selected IP Traffic Offload SIPTO*, *Local IP Access* (LIPA), y *Flow Mobility And Seamless Offload* (IFOM). El 3GPP en la Release 10 propone SIPTO como mecanismo para liberar tráfico de la macrocelda, independizando el tráfico de la macrocelda hacía una nueva área específica dentro de la red de acceso. Existe además la posibilidad de intercambiar la información directamente dentro de un área, formando una red privada que agrupa varios MNs. Área que se enlaza con el eNB a través de un *Local Gateway* (L-PGW) para formar una estructura de small-cell. (Ver figura 10-b). El MME selecciona la pasarela más apropiada para el tráfico dependiendo de la proximidad del MN y los parámetros almacenados en el HSS. El tráfico SIPTO no se encamina hacia el core de red, la movilidad en SIPTO en la macrocelda y entre red formada por la pasarela de acceso L-PGW y el eNB se gestiona a través de la interfaz S5, son los mismos procedimientos de movilidad definidos en la release 10.

Por otro lado, LIPA es un mecanismo por el cual un MN conectado al eNB es capaz mantener la misma dirección IP durante el movimiento del MN cuando el MN se mueve entre eNBs que forman un grupo (Ver figura 10a). En LIPA el L-GW actúa como punto de anclaje del MN durante el movimiento del MN en las redes de acceso. La pasarela contiene un buffer capaz de almacenar paquetes cuando el MN cambia de eNB. El concepto de LIPA se introdujo en la Release 10, aunque funcionalmente operativo en la Release 11, en la que se realiza una separación entre el L-GW y el eNB y donde se soporta la movilidad de flujos. La activación y desactivación del contexto LIPA PDN la lleva a cabo el MME/SGSN, quien decide que conexiones necesitan ser traspasadas al L-GW. (TR 23.859). Cuando el MN solicita una portadora en LIPA se establece una sesión PDN que termina en el L-GW, en ese momento se le asigna una dirección IP al MN, dirección que pertenece a la red local local formada por los eNBs y el L-GW. Todos los paquetes que llegan se guardan en el buffer del L-GW que los encamina hacia el eNB donde se encuentra el MN.

Finalmente en la Release 10 se introduce el concepto IFOM. Es un mecanismo para transportar flujos de tráfico a través de diferentes redes de acceso dentro de un mismo contexto PDN. Si hay que mantener varios contextos PDN se especifica *Multi Access PDN Connectivity* (MAPCON). Es el mecanismo más eficiente de todos, y analizado a continuación.

El 3GPP en el *Technical Specification* 23.261 especifica el modo de soportar varios flujos en un mismo contexto PDN utilizando diferentes accesos de red de forma simultánea dentro del EPS, ya sea en redes 3GPP, o non-3GPP. El 3GPP presenta dos alternativas para accesos WLAN, una solución basada en host utilizando DSMIPv6 sobre el interfaz S2c, y una solución basada en red utilizando PMIP sobre los interfaces 2Sa, 2Sb.

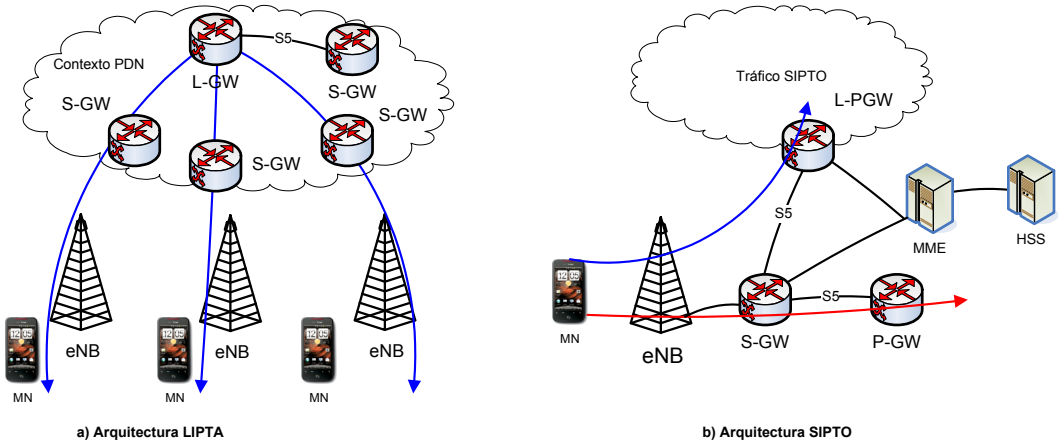

Figura 10. Arquitecturas Traffic Offload TR 23.859

3.1.1. Gestión de flujos móviles con IFOM

Proveer movilidad en redes convergentes IP introduce nuevos desafíos, uno de ellos surge como consecuencia de la imposibilidad de integrar el protocolo MIP con el sistema IMS. La arquitectura IMS ofrece movilidad para servicios sensibles al retardo, dentro de dominios administrados a través de la señalización proporcionada por el protocolo SIP que identificar los extremos de una conexión para gestionar la movilidad del MN a través de la arquitectura IMS.

El enfoque no es óptimo en términos de latencia. Tal circunstancia obliga a introducir nuevos mecanismos capaces de encontrar la ruta óptima entre los dos extremos. Un mecanismo consiste en gestionar *flujos de tráfico,* alternativa eficaz que mejora la QoS [18] y reduce la latencia. La gestión de flujos permite optimizar el throughput del servicio en términos de tasa de bit. La solución está basada en principios DSMIPv6, permite mover flujos dentro de las redes de acceso. El mecanismo fue estandarizado por el IETF y adoptado por el 3GPP, para transportar flujos dentro de un contexto PDN utilizando los diferentes interfaces del MN. Aunque en este sentido se deben modificar, y extender los protocolos MIPv6/PMIPv6. El mecanismo permite que un usuario pueda beneficiarse de coberturas de WLANs cercanas, para mejorar la QoE, eso se consigue descargando el tráfico de la macrocelda, también se lleva a cabo el traspaso en el caso de que aumente la posibilidad de interrumpir el servicio. La solución no introduce duplicación de paquetes, ni perdidas de paquetes debido al coste en términos de tiempo en adquirir la nueva configuración en la capa de red por parte del MN.

Un flujo móvil se define como un servicio IP, o VoD soportado por un dispositivo con varios por interfaces. Es posible asignar a cada flujo diferentes políticas QoS predefinidas, los flujos pueden sumistrarse a través de diferentes accesos a red, cada flujo puede proveer diferentes tipos de servicio, los flujos pueden moverse entre diferentes accesos si aumenta la probabilidad de pérdida de conectividad. Con este fin, el 3GPP y el IETF, adoptan y proponen sus soluciones basadas en PMIPv6 extendido nuevas extensiones del protocolo. Las soluciones propuestas adquieren un impacto relevante dentro de las nuevas arquitecturas móviles, donde es posible intercambiar sesiones 4G, con sesiones WLAN. Aplicar QoS a un flujo es más sencillo, es posible mapear con mayor granularidad las CoS en función del perfil del servicio a un flujo. Ello obliga a una reflexión, la QoS se especifica en términos de parámetros de red ó por el contrario la QoS son los requisitos que necesita un determinado consumidor de un servicio, parece evidente que son los parámetros que necesita el servicio y deben proporcionarse por la infraestructura a través de la señalización.

Por lo tanto, es posible mapear preferencias de usuario a través de políticas basadas en parámetros de red, y caracterizadas en términos de QoS. Por ejemplo, parece razonable que un flujo para un tráfico VoIP sea suministrado por una red 4G para aprovechar mejor gestión de la QoS y el tráfico best-effort sea suministrado por una conexión WLAN sin gestión de QoS.

En este nuevo escenario y dentro de entornos multi-interfaz aparecen en la Release 10 aparecen los flujos móviles IFOM que pueden ser inicializados por las entidades centrales como el HA, o el LMA.

De ese modo, el aprovisionamiento de recursos se realiza según las necesidades del servicio. Es posible mover el flujo de datos entre las interfaces del MN, en el caso de que existan retardos, o revocarlo en caso de que el flujo no cumpla con las expectativas de usuario. La idea se contempla en los documentos RFC 5555, RFC 5648, y RFC 6089. El documento describe las extensiones MIPv6 que permiten enlazar a un MN uno o varios flujos hacia su CoA con soporte multi-interfaz. Las extensiones permiten registrar varias CoA de un MN en el HA, sin más que extendender los mensajes BU definidos en MIPv6 para que sean capaces de soportar el identificador del interfaz. Se identifica el interfaz mediante *Binding Identifiers* (BID) y posteriormente se registrar una entrada en la *Binding Caché* (BC) mapeando el BID.

El procedimiento en MIPv6 es el siguiente. El HA añade en la BC una entrada por cada una de las CoA del MN. Posteriormente registra los flujos de sus CoA, mediante las extensiones *Flow Binding*, que permiten a un MN enlazar uno o varios flujos a determinada CoA. De ese modo el MN puede señalizar el como enviar información a un CN por un determinado interfaz. Las especificaciones flow bindings establecen el como asociar al MN un derminado flujo identificado mediante FID con una CoA-BID. Los bindings o enlaces entre un flujo binario y la entrada de la cache están almacenados en zonas de memoria separadas que se enlazan a través del campo BID. La *flow binding* incluye el identificador de flujo FID, el traffic selector es el grupo de filtros que caracterizan el flujo binario (RFC 6089), el FID-PRI que es la prioridad, y el BID asociado. La binding cache contiene la estructura extendida que permite almacenar el identificador de CoA mediante el BID. (Ver figura 11).

Flow bindings			
FID-PRI	FID	Traffic Selector	BID
10	FID1	scrAddr=CN1	BID1
20	FID2	scrAddr=CN2	BID2
30	FID3	TCP	BID3

Binding cache			
HoA	BID	CoA	BID-PRI
PrefijoH::MN	BID1	Prefijo1::MN	5
PrefijoH::MN	BID2	Prefijo2::MN	10
PrefijoH::MN	BID3	Prefijo1::MN	20

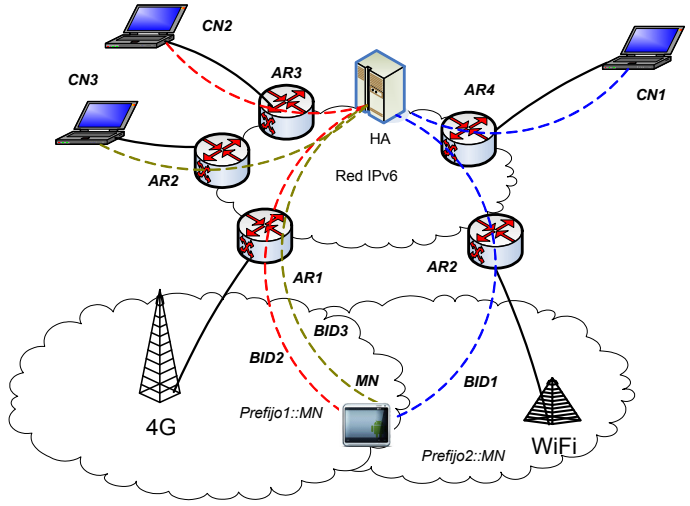

Figura 11. Extensión MIPv6 para realizar un handovers de flujos de datos

El procedimiento para el protocolo PMIP es similar a MIP. En un contexto de sesión PMIP, cuando el MN modifica su posición todos los flujos asociados al MN se mueven hacia el nuevo MAG. El planteamiento choca con la idea original IFOM propuesta por el 3GPP que describe como en el procedimiento de handover solamente deben estar involucrados aquellos flujos que realmente lo necesiten desde el punto de vista del servicio en términos QoE. Por lo tanto, el MAG debe ser capaz de reencaminar solamente determinados prefijos IPv6, incluso si estos prefijos son suministrados por MAGs diferentes.

En el diseño original del protocolo PMIPv6, se definen prefijos únicos y sesiones diferentes para cada una de sus interfaces. El protocolo no especifica como el MAG puede encaminar el tráfico, utilizando solamente determinados prefijos IPv6. Tal circunstancia impide la posibilidad de separar el tráfico a través de diferentes interfaces y diferentes MAGs.

De ese modo, se extiende el protocolo para que sea capaz de soportar movilidad basada en flujos, solamente a partir de la información almacenada en la memoria. Ser capaz de recibir y enviar tráfico, independientemente del interfaz. Cualquier extensión, ó modificación del protocolo debe ser independiente del MN, por estar basado en PMIP.

En PMIPv6 existen dos modos de asignar identificadores de red a los interfaces del MN. En primer lugar el MAG asigna un conjunto de prefijos únicos IPv6 por interfaz, el LMA actualiza la cache para cada una de las entradas a partir de los identificadores de red asignados al interfaz del MN, la actualización se lleva a cabo mediante señalización procedente del MAG. En cuanto al segundo mecanismo de configuración, existe la posibilidad de asignar un mismo prefijo IPv6

para cada interfaz, aunque en este caso cada interfaz se autoconfigurara con una dirección global *unicast* única. En ambos casos, un handover implica que un único MAG administre los traspasos de los flujos desde un interfaz hacia el otro interfaz.

El nuevo mecanismo extiende el protocolo PMIPv6 para encaminar solamente determinados flujos. En este sentido, para que la solución sea independiente del terminal, la solución pasa por crear un interfaz lógico (*Logical Interface* (LIF)), que oculte los interfaces físicos [19], y permita a las capas superiores ver solamente un único interfaz. De modo que para los protocolos de capas inferiores serán transparentes a los protocolos de capas superiores. El mecanismo permite utilizar varias redes de acceso con tecnología diferente como LTE, 802.16, y IEEE 802.11 [19], siendo transparente al dispositivo.

El mecanismo permite tener sesiones diferentes en el LMA en función del prefijo de red IPv6 asignado. Supongamos que el interfaz 1 se une al MAG1 con soporte 4G y que el interfaz 2 se une al MAG2 con soporte WLAN. El LMA debe ser capaz de mantener las dos sesiones diferentes a través de sendas entradas en su binding cache. La configuración de los interfaces en el MN se realiza mediante prefijos IPv6, prefijos global unicast diferentes. Para ello, se utilizan mensajes de señalización de tipo PBU/PBA, encargados de señalizar la sesión entre MAG-LMA. La solución PMIP es escalable y eficiente en términos de latencia y throughput.

La figura 12 representa un escenario en el que se lleva a cabo un handover para un flujo de tráfico binario. El ejemplo ilustra un MN con dos interfaces de red, el interfaz if1 enlazado con la MAG1, y el interfaz if2 enlazado con la MAG2 cada interfaz recibe un prefijo IPv6 diferente. El MN recibe el flujo X a través del interfaz if1, y recibe el flujo Y a través del interfaz if2. Los dos flujos son diseccionados a través de su interfaz virtual. El LMA mantiene una estructura de estado hacia que MAG se encaminan los flujos. De ese modo el LMA posee en su caché una estructura que representa el estado de la movilidad que mapean los diferentes flujos hacia los MAG. El mapeo se realiza según la recomendación RFC 6088.

En un momento determinado, y debido a un perfil de trafico asociado al MN, el LMA decide mover un flujo "Y" desde el MAG2 hacia el MAG1. La decisión se basa en perfiles móviles de usuario, debido por ejemplo a que existe una fuerte congestión en la red. El LMA señaliza el procedimiento al MAG1 para que este actualice su tabla de rutas, señaliza que flujo va a traspasar al MAG1. Posteriormente el LMA modifica su flow binding para mapear la ruta al nuevo flujo, en ese momento empieza a enviar tráfico a través del MAG1.

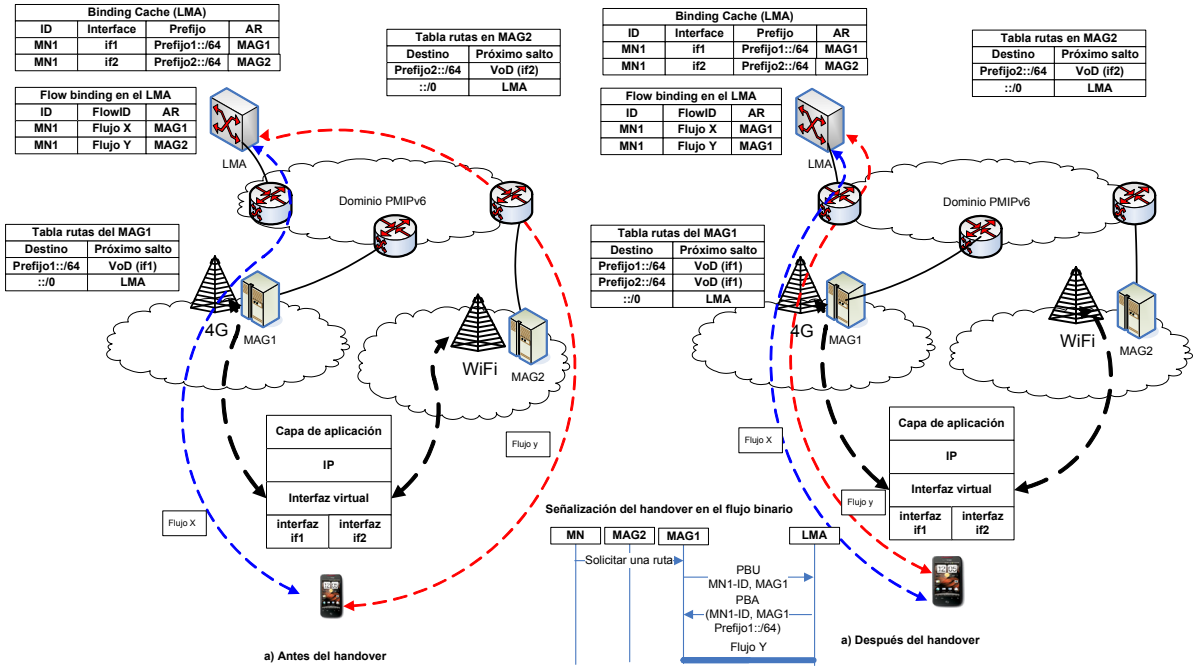

Figura 12. Procedimiento flow mobility en PMIPv6

3.1.2. Gestión de handovers transparentes con ANDSF

Independientemente del escenario de análisis el MN obtiene una dirección IP en cada una de las redes de acceso que visita, el cambio de dirección en capa de red no debe afectar a la sesión en curso, la señalización debe ser ajena al MN. Por otro lado, la infraestructura debe ser un agente ubicuo para el trafico IP. En este sentido, las soluciones que se proponen deben garantizar que la infraestructura proporcione la ruta óptima con los mínimos recursos de red posibles. A partir de los requisitos funcionales anteriores, se presenta una solución capaz de conseguir que la solución de *handover* sea ubicua para la tecnología y la red de acceso al MN.

El EPC integra accesos 3GPP y accesos non-3GPP. De ese modo, es posible realizar *handovers* intra-area o inter-area. Los *handovers* intra-area se realizan entre un mismo dominio tecnológico. Los handovers inter-area son *handovers* verticales llevados a cabo entre tecnologías non-3GPP y 3GPP. Durante el handover es necesario seguir manteniendo la continuidad en la sesión. En el caso de que la movilidad se gestione en capa 2, la movilidad puede ser vista como una única tecnología de radio acceso. Por el contrario, si la movilidad es gestionada en capa 3, es necesario que el MN configure diferentes direcciones IP fijas durante el recorrido en las diferentes subredes que visita en MN.

La solución propuesta por el 3GPP dentro del EPS consiste en definir una entidad *Access Network Discovery and Selection Function* (ANDSF), entidad que proporciona información para seleccionar la mejor tecnología de acceso que permita descargar el tráfico IP, o mejorar la QoS del servicio a partir de requisitos previamente definidos. De ese modo, las reglas preconfiguradas en el ANDSF permiten intercambiar accesos 3GPP o accesos WLAN, la movilidad IP se gestiona a través de los protocolos MIP, en el caso de soluciones basadas en host, o PMIP para soluciones basadas en red.

La solución esta estandarizada por el 3GPP y los procedimientos documentados en el *technical specification* TS 23.402. Procedimientos que ayudan a seleccionar el mejor acceso a red para descargar el tráfico de la macrocelda, preferiblemente accesos WLAN. De ese modo el operador descarga cierto tráfico hacia una red de acceso determinada. Para ello, el ANDSF proporciona información de las diferentes redes de acceso próximas al MN. Los handovers se realizan dinámicamente a partir de perfiles de tráfico individuales y perfiles de movilidad de usuario. Tal circunstancia permite buscar a iniciativas para centralizar la señalización o incluso llevar la señalización a instancias del *cloud computing,* situación que discute actualmente.

En definitiva, se plantean soluciones que facilitan la toma de decisiones por parte del MN para optimizar el rendimiento de los servicios que le suministra la infraestructura, especialmente aquellos sensibles al retardo, jitter, y pérdida de paquetes.

La Release 8 define los procedimientos para que un MN pueda conectarse al EPC utilizando accesos 3GPP, o non-3GPP como WiFi ya sean trusted o untrusted. Un acceso untrusted es aquel en el que un MN se conecta a una WLAN de un proveedor diferente al suyo, o en el caso de que el subscriptor utilice una conexión a Internet para conectarse con su operador. En este caso, las especificaciones permiten establecer un túnel IPSec/IKEv2 seguro hacia la entidad Evolved Packet Data Gateway (ePDG) antes de que el tráfico sea encaminado al núcleo de red del operador. Por el contrario, en el caso de que el usuario se conecte a la red WiFi del operador, se lleva a cabo un acceso trusted, no siendo necesario crear un túnel seguro, tampoco es necesario en ese caso la entidad ePDG, debido a que el MN se ancla directamente al P-GW.

El ANDSF mejora la gestión para accesos trusted y untrusted definidos por el 3GPP. Actúa como un servidor dentro del EPC almacenando información de red, por ejemplo información de la ubicación de los APs con el objetivo de optimizar el tráfico de red, para que el MN pueda seleccionar las mejores redes de acceso en un instante de tiempo determinado. La entidad contiene información para que el MN puede interrogar, y de ese modo obtener un listado de redes candidatas para realizar el *handover* [20].

Actualmente el 3GPP implementa el ANSDF como solucion basada en host untilizando el protocolo DSMIPv6 a través del interfaz 2Sc. El MN interroga al ANSDF a través del interfaz S14. El ANSDF es capaz de traspasar flujos entre las redes de acceso.

El ANSDF proporciona información útil que sirve para seleccionar el mejor tipo de acceso según preferencias de usuario predefinidas. También muestra un listado de redes de acceso disponibles proporcionando información relevante que ayude al MN a tomar las mejores decisiones en el handover. Y por último, proporcionar determinadas reglas que ayuden al MN a seleccionar el interfaz hacia el que mover el flujo de tráfico.

La Release 10 proporciona los procedimientos para que el MN pueda conectarse simultáneamente a redes de acceso diferentes. El procedimiento tiene tres modos de operación. El primero se denomina *Multi-Access PDN Connectivity* (MAPCON) no es más que la capacidad del MN para poder fijar, y establecer diferentes redes de acceso en diferentes contextos PDN. En cuanto al segundo se denomina, *IP Flow Mobility* (IFOM) se refiere a los mecanismos para trasladar flujos entre diferentes interfaces sin perder la conectividad. Finalmente, el *Non-Seamless Offload* que no es más que la capacidad del MN para elegir el flujo IP que le proporciona las mejores prestaciones en términos de QoE.

Actualmente las especificaciones se encuentran en un estado tecnológico inmaduro, los desafíos que presentan son numerosos, uno de ellos es la duración de las baterías, especialmente en áreas con baja densidad de APs, debido a que el MN intenta descubrir redes periódicamente con el consiguiente gasto en el consumo de la batería. Se debieran mejorar los algoritmos que obtienen información de usuario e información de red. En cuanto al segundo inconveniente destacar la calidad de servicio, el MN debe ser capaz de seleccionar de modo automático el mejor acceso sin tener en cuenta al usuario, y por último la capacidad para localizar el MN en entornos de interiores, actualmente se lleva a cabo con técnicas de posicionamiento *Global Positioning System* (GPS), técnicas poco precisas en interiores. La capacidad de identificar la localización del MN es crucial para los sistemas ANDSF. Todas las políticas de aprovisionamiento de QoS están basadas en la posición de la localización del MN, posición obtenida a través de los sensores del MN, como el acelerómetro o el GPS. El acelerómetro consume menos energía es un mecanismo utilizado por diversas aplicaciones, donde la duración de la batería es un factor determinante.

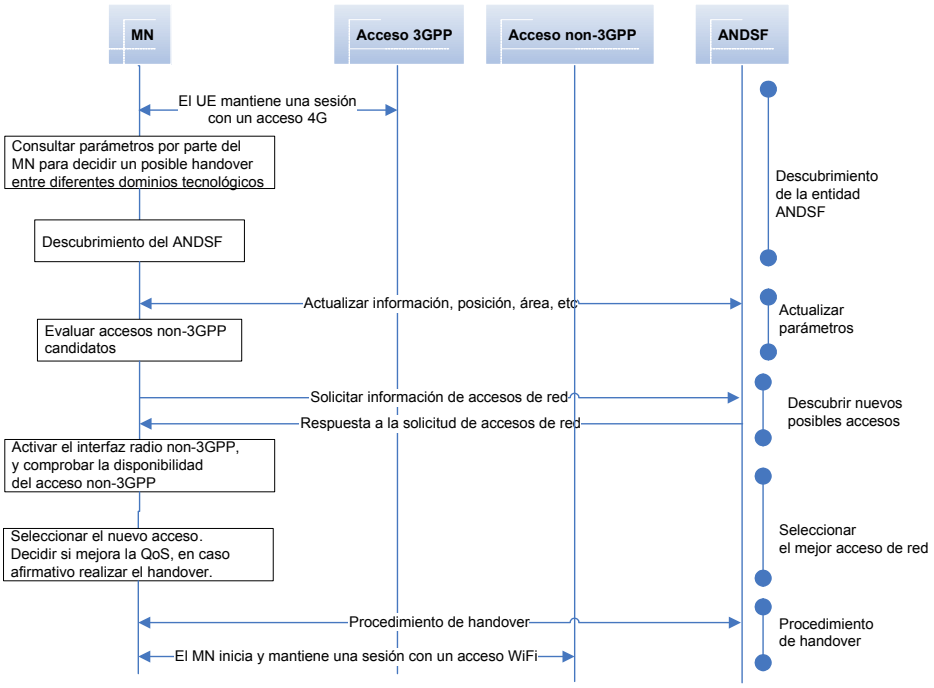

Figura 13. Señalización de un handover non-3GP a través de ANDSF

3.2. Propuestas para disminuir el retardo E2E con protocolos IETF

Las propuestas en arquitecturas 4G estudiadas, por si solas no proporcionan todos los servicios necesarios en la red de acceso de forma integrada, esto es debido a que la mayoría de las investigaciones se orientan hacia la red troncal y los servicios de usuario, o en algún servicio específico de la red de acceso. Por lo tanto, es necesario seguir analizando propuestas que permitan analizar y descubrir la complejidad que conlleva suministrar servicios por parte de la red de acceso, a través de distintas tecnologías de acceso, utilizando para ello una arquitectura de red en el backhaul común.

3.2.1. Contenedores de flujos IPv6 en el EPS

Actualmente solamente los servicios VoIP soportan movilidad real en el EPS, esto es debido a que solamente se asigna un gateway local móvil a un servicio VoIP, debido a los problemas que ocasiona el retardo durante el movimiento del MN. En este contexto, se centran las investigaciones actuales, como encontrar nuevos mecanismos capaces de anclar los servicios sensibles al retardo, en particular los servicios de video. Con ese fin el trabajo analiza el campo etiqueta de flujo dentro de la cabecera IPv6, campo capaz de mapear el tráfico a un sistema portador tanto en las redes de acceso como en el backbone [21].

La implementación del protocolo IPv6 presenta algunos inconvenientes, uno de ellos es el excesivo overhead de red necesario para señalizar las soluciones tuneling en la traducción de direcciones, con el objetivo de proporcionar una pila dual, compatibilidad IPv4 y IPv6. Entre los aspectos positivos de la implementación del protocolo están el soporte QoS a través de los campos de cabecera y la ausencia de NAT para reenviar tráfico [22]. Una ventaja es la posibilidad de agregar varios prefijos IPv6 en cada uno de los interfaces del MN, prefijos anunciados automáticamente a través de mecanismos de descubrimiento de ruta y descubrimiento de vecinos.

Actualmente, los mecanismos de gestión de QoS dentro del EPS están basados en servicios de portadoras en las redes de acceso, y técnicas de gestión de recursos como Diffserv, MPLS, y RSVP en el backhaul IP.

Las redes de nueva generación particularmente redes 4G fueron diseñadas para ser redes IP, circunstancia que permite introducir de modo nativo el protocolo IPv6, el protocolo está disponible en los diferentes dominios administrados que componen un servicio E2E. Así pues, es posible introducir IPv6 en las redes de acceso, en el core de red, en los servicios portadores IP sobre el interfaz S1, en las interfaces S5/S8, y en los servicios externos proporcionados por los usuarios finales más allá de las pasarelas P-GW.

Para hacer agnóstica la QoS de la tecnología se definen los servicios portadores. Servicios que actualmente contienen varios flujos de tráfico de un mismo contexto PDN. De ese modo, introducir el protocolo IPv6 en el EPS impacta directamente en la QoS del servicio portador móvil, porque es posible definir el nivel de granularidad de la QoS en un flujo.

El concepto Service Data Flows (SDF) forma parte de la estrategia del 3GPP para implementar mecanismos de QoS, permite asignar flujos de tráfico a portadoras a partir de selectores de tráfico. El 3GPP define un servicio portador para transportar datos en el plano de usuario entre nodos de red, y proveer de forma fácil QoS a los flujos asociados al servicio portador. Varios SDF se pueden multiplexar en una misma portadora dentro del EPS. Es fácil darse cuenta que el mecanismo es poco flexible, debido a que la QoS actualmente solamente se puede aplicar a un servicio portador, y este puede contener varios flujos multiplexados.

El despliegue de servicios sensibles al retardo como VoD, juegos online, obliga a introducir nuevos mecanismos de gestión de recursos. Un aspecto importante es el tamaño de cabecera IPv6, el doble que la cabecera IPv4, situación que afecta notablemente a servicios caracterizados a través de paquetes pequeños como un servicio VoIP. De ese modo, la relación

IPv6/QoS impacta negativamente en términos QoS dentro de los mecanismos de portadoras, y en los mecanismos de aprovisionamiento de recursos. Para solucionar el problema se utilizan en este sentido, mecanismos de compresión de cabecera. El objetivo es evitar enviar campos que no cambian entre paquetes consecutivos, la cabecera viaja comprimida, por ejemplo una cabecera en IPv6 de 60 bytes, puede comprimirse en 3 bytes. Aunque el mecanismo introduce latencias en la compresión, y descompresión de la cabecera, posteriormente veremos como eliminar la latencia.

Para mejorar el diseño del sistema se deben tener en cuenta factores críticos que impactan directamente en la QoS de redes móviles, en primer lugar el rendimiento del canal radio. En segundo lugar la capacidad en términos de throughput del enlace E2E. También el diseño de la arquitectura de red, la capacidad de minimizar el retardo máximo y conseguir la mayor tasa de bit posible en servicios E2E, y finalmente las características del servicio. De ese modo, los mecanismos para mejorar la QoS afectan a cada una de las capas y elementos de red que proveen el servicio E2E.

En ese sentido, el 3GPP define y provee mecanismos QoS E2E en el EPS. El EPS proporciona conectividad con redes externas a través del P-GW. El P-GW soporta *Traffic Flow Agregates* (TFT), una tupla capaz de identificar el flujo binario que se crea cuando se crea un sistema portador en el EPS. De ese modo, el UE y P-GW manejan el tráfico con la información contenida en la tupla. El MN, S-GW, y el P-GW utilizan los filtros TFT para mapear flujos IP a portadoras. Cuando se establece una portadora se crean los filtros TFT circunstancia que permite gestionar de ese modo el tráfico en cada nodo involucrado.

El protocolo IPv6 impacta directamente en los mecanismos TFT ya que los filtros TFT contienen información de cabecera del protocolo como campo próxima cabecera, campo clase de tráfico, y campo etiqueta de flujo (RFC 3697). Por ejemplo, un posible filtro puede estar formado por las direcciones origen y destino, la clase de servicio, y el identificador de flujo. De este modo, se consigue un ajuste más fino para implementar políticas QoS. Actualmente, cada portadora contiene asignados varios flujos de tráfico asociados al servicio dentro de un mismo contexto PDP, es difícil de ese modo aplicar QoS. Por lo tanto, es complicado ajustar con cierto nivel de detalle la QoS asociada a un flujo. Por ejemplo, diferentes pipelines asociados a un servicio http viajan en una misma portadora.

Como implementar una estrategia eficaz todavía es un tema de estudio en la comunidad científica (RFC 6438), (RFC 6294). El gran avance de IPv6 se orienta en ese sentido, clasificar los paquetes mediante tuplas, y conseguir separar el flujo del contenedor. Algunos autores proponen la arquitectura *IPv6 Label Switching Architecture* (6LSA) como mecanismo que agrupa las dos características anteriores, siendo capaz de garantizar QoS y encaminar paquetes, todos los paquetes con la misma etiqueta reciben el mismo tratamiento en el nodo.

El campo etiqueta de flujo tiene las siguientes propiedades, la no encriptación en caso de usar IPSec. Está presente en todos los fragmentos del paquete, circunstancia que le convierte a ser un sucesor de la tecnología MPLS, cualquier paquete que no pertenece al flujo binario almacena un cero. En cada paquete existe una tupla de tres campos que identifica unívocamente a ese campo, el paquete recibe un tratamiento especial en ese campo, siempre que el nodo soporte la característica. Es posible aleatorizar, es un mecanismo perfecto para securizar el flujo. (RFC 6294).

En la RFC 6294 se especifica como se puede usar el campo etiqueta de flujo para encaminar paquetes, en este sentido el aumento del tráfico multimedia requiere que se le de todavía mayor importancia, y que paquetes conmuten por etiqueta, en lugar de encaminarse del modo tradicional. Esta distinción en los flujos permite que los routers reaccionen mejor en caso de congestión. Actualmente, los mecanismos de aprovisionamiento de QoS se basan en mecanismos Diffserv, o mecanismos RSVP. El campo etiqueta de flujo permite adaptar el flujo

a mecanismos QoS Diffserv para ello mapea el flujo a través de la cola Expedited Forwarding (EF), solución que permite escalar fácilmente el tráfico.

Por lo tanto, el análisis demuestra como el protocolo IPv6 optimiza latencias. Ayuda a la mejora de la QoS en los interfaces S5, S8, S2a, S2b, y S2c debido a la posibilidad de gestionar flujos de modo nativo. Utilizar el protocolo IPv6 ayuda a independizar los flujos del contenedor, y asignar la misma clase de servicio durante toda la ruta. Importancia que adquieren los interfaces basados en PMIPv6/DSMIPv6. Esto es importante debido a que PMIP no soporta el mapeo entre portadoras. El servicio de portadoras se diseño en redes 3GPP para proveer QoS con técnicas de priorización de recursos a través de CoS. Los flujos PMIP, siempre viajan encapsulados en el mismo túnel a través de los interfaces S5/S8 desde el LMA hacia el MAG.

También debemos tener en cuenta en la optimización de latencias la ubicación del P-GW, elemento que actúa de LMA en PMIPv6, y HA en MIPv6. Optimizar la ubicación de las entidades consigue mejorar significativamente las latencias. Debemos tener en cuenta que los operadores utilizan pocos nodos para enlazar el tráfico con redes externas, y que a menudo existe gran distancia entre nodos. La probabilidad de que un MN cambie de P-GW durante su movimiento es nula, por lo tanto ubicar el LMA en el P-GW no es una solución óptima. En cambio, si hacemos que el LMA alojado en el P-GW, proporcione servicio a áreas más pequeñas o tener una funcionalidad más especifica, conseguimos mejorar la latencia.

3.2.2. Movilidad dinámica en HetNet

El elevado tráfico a cursar dificulta su caracterización. Es difícil señalizar los flujos salto a salto en cada nodo de red, y de ese modo aprovisionar recursos. Actualmente, la ineficacia de los mecanismos implica ineficiencias en la infraestructura sobre la que se transporta el tráfico. Situación especialmente crítica en las redes de acceso donde se cursa la mayoría del tráfico. Tal situación provoca una transformación y reorganización de los elementos funcionales en las redes de acceso.

En este apartado se enumeran mecanismos para optimizar la latencia, mecanismos que surgen en un contexto de movilidad dinámica. Mecanismos que proponen flexibilizar el diseño de las arquitecturas de red actuales. Dos son las principales estrategias de futuro orientadas hacía la optimización de la latencia E2E. La primera optimizar el diseño de coberturas para que sean escalables, conseguir alta disponibilidad y escalabilidad. En cuanto a la segunda estrategia está basada en la optimización de la utilización del espectro, limitado y caro por naturaleza, circunstancia que impide que el ancho de banda pueda incrementarse fácilmente.

El paradigma para optimizar el tráfico consiste en adaptar el servicio a la infraestructura, y no la infraestructura al servicio. Los operadores evolucionan en este sentido, por ejemplo buscando técnicas capaces de optimizar el espectro, propuesta LTE. Desplegando nodos de baja potencia, propuesta femtoceldas, y picoceldas. Finalmente la posibilidad de cursar el tráfico de la celda hacía tecnologías alternativas, propuesta WiFi.

De este modo, la movilidad se convierte en una característica importante para los servicios sensibles al retardo, especialmente en arquitecturas HetNet de femtoceldas, caracterizadas por ser estructuras capaces de operar bajo espectro licenciado y no licenciado, en este contexto los servicios *seamless* añaden nuevos problemas que deben ser resueltos.

Las soluciones orientadas a la gestión de la movilidad IP analizadas en el trabajo, están basadas en arquitecturas jerárquicas y centralizadas. Arquitecturas que utilizan los MN durante el anclaje en la red externa para señalizar su movimiento y reenviar tráfico. En redes 4G y redes HetNet, los puntos de anclaje siguen siendo entidades centrales [23]. Como consecuencia, aparecen ciertos problemas de escalabilidad, por ejemplo añadir nuevos MNs sobrecarga la infraestructura, esto es debido a que el tráfico generado siempre viajará primero a

la red origen para encaminarse después hacía la red actual, introduciendo de ese modo un retraso, sobrecarga y señalización. Además, el mecanismo produce una gran sobrecarga computacional en los nodos centrales, debido a que deben procesar un gran volumen de tráfico. En numerosas ocasiones el MN, y el CN están relativamente cerca pero aún así la información viaja hacía un nodo central. De ese modo se consiguen rutas no optimizadas. También hay que tener en cuenta que parte del tráfico no siempre es necesario que fluya hacia el core de red, debido a que se genera y consume en las redes de acceso.

El problema se amplifica debido a existencia de pocos puntos de anclaje en la red, los pocos puntos de anclaje gestionan el tráfico de millones de MN, la solución es poco tolerante a errores. Actualmente, casi todas las mejoras pasan por estandarizar nuevas extensiones MIPv6, extensiones que conservan el diseño inicial centralizado y jerárquico. Extensiones desarrolladas para mejorar, y mantener controlada la latencia. Su diseño, hereda el mismo problema, soluciones poco escalables, rutas poco óptimas, excesivo *overhead* de red, y una elevada pérdida de paquetes en servicios E2E, de ese modo las soluciones hacen un uso ineficiente de los recursos de red especialmente en el backhaul y en las redes de acceso.

En este sentido, surgen propuestas que tratan el problema de la movilidad desde una perspectiva dinámica de la movilidad. La idea subyacente consiste en reubicar los puntos de anclaje en la red, y por tanto reubicar las funciones encargadas de gestionar la movilidad en sistemas inalámbricos. La solución consigue de ese modo esquemas y estructuras distribuidas. La propuesta permite colocar los puntos de anclaje en diferentes ubicaciones dentro de la arquitectura, para que estén tan cerca del usuario como sea posible. La solución se denominada *smart-mobility*, no es otra cosa que proveer movilidad dentro del área donde a ciencia cierta sabemos que se desplazará el MN.

La solución consigue escalar, significa que añadir más MNs no es un problema, solamente debemos añadir más puntos de anclaje. Propuestas realizadas por el IETF [24] [25].

Bajo ese punto de vista, se denomina *Distributed Mobility Management* (DMM) la distribución de los puntos de anclaje en la red. Cada red de acceso posee una entidad capaz gestionar el tráfico hacia dentro del backbone o hacia las redes de acceso continuas, independientemente del resto de elementos de la infraestructura. El tráfico solamente se encamina al backbone si fuera necesario. La movilidad se gestiona nodo a nodo. El MN se enlaza al punto de anclaje cuando alcance la nueva ubicación en la red externa de acceso.

En este sentido, la solución distribuida propuesta por algunos autores debe cumplir los requerimientos que propone el IETF [26]. El organismo, establece los requisitos que debe cumplir cualquier solución distribuida. De este modo, cualquier propuesta que se proponga para gestionar la movilidad debe enfocarse desde una aproximación basada en host, una aproximación basada en red, y por último una aproximación basada en el encaminamiento dinámico de rutas. A continuación proponemos, y desarrollamos los tres enfoques de la solución DMM.

En la aproximación basada en host. El MN encapsula o desencapsula el tráfico que el AR le suministra. La solución fija uno o varios puntos de anclaje par el MN, en este caso el HA se ubica en cada uno de los routers de accesos. De ese modo el MN no utiliza solamente un prefijo IPv6, utiliza uno diferente que autoconfigura en cada router de acceso de la red visitada. Esto significa que el MN puede inicializar nuevas sesiones en cada nuevo HA, también significa que debe mantener las sesiones en curso en cada nuevo HA. Las sesiones en curso se aseguran con túneles bidirecionales entre el MN y los HAs. Para ello el MN enlaza con procedimientos BU su nueva HoA en cada uno de los nuevos HA. El MN también debe ser capaz de manejar múltiples direcciones IPv6 simultáneamente.

Por ejemplo, un MN1 se une con el HA/AR1 y configura su IPv6 HoA1 para intercambiar paquetes con CN1. El movimiento del MN1 obliga a enlazarse con el HA/AR4, de ese modo

configura una nueva IPv6 HoA4. El MN utiliza la HoA4 para intercambiar paquetes con CN2, la continuidad de sesión con CN1 queda asegurada con un túnel entre MN1 y AH/AR1.

La solución DMM basada en MIPv6 soporta optimización de ruta. El MN puede establecer una sesión con el CN directamente. El MN negocia con el CN su nueva CoA a través de mensajes BU. Una vez confirmados se lleva a cabo el túnel bidireccional entre el MN y el CN.

Solucion basada en red. (Ver figura 14). El MN no interviene en la señalización del movimiento. La solución también ubica la gestión de la movilidad en los routers de acceso en cada acceso de red. El router de acceso desarrolla funcionalidades duales LMA/MAG. Son entidades lógicas que actúan como puntos de anclaje para los MNs. Existen drafts en el IETF que denominan al punto de anclaje *Distributed Gateway* (D-GW). Es el nombre que adoptamos en el trabajo, la solución utiliza el protocolo PMIPv6. La idea es tener varios D-GW que gestionen únicamente pequeñas ubicaciones, de ese modo son los responsables de pocos subscriptores.

La solución DMM que plantean algunos autores es similar a la solución anterior basada en host, el MN adquiere direcciones CoA virtuales por cada uno de los routers de acceso visitados. Direcciones que son asignadas dinámicamente por los D-GW visitados, entidades lógicas dentro del AR. Los AR se encargan de encaminar el tráfico hacia el nuevo AR mediante un túnel. Cada nuevo AR asigna un prefijo IPv6 al MN para que autoconfigure su dirección. El MN puede empezar una nueva sesión utilizando los nuevos prefijos IPv6 asignados. El tráfico en curso, es transportado por túneles entre los diferentes AR en las redes de acceso.

La figura 14 representa un posible escenario en el que el MN configura un prefijo IPv6 en el router de acceso AR1 (Prefijo1:MN/64) y empieza el envió de paquetes utilizando esa configuración. En un momento determinado se desplaza y se enlaza con el router de acceso AR2, el router suministra al MN el identificador de red Prefijo2::/64. De ese modo, el MN es capaz de empezar nuevas sesiones utilizando el prefijo asignado, por ejemplo puede empezar una sesión hacia CN2, y también puede mantener las sesiones en curso, por ejemplo hacia CN1 a través del AR2. Las sesiones en curso se mantienen mediante la creación de un túnel bidireccional entre el AR1 y AR2.

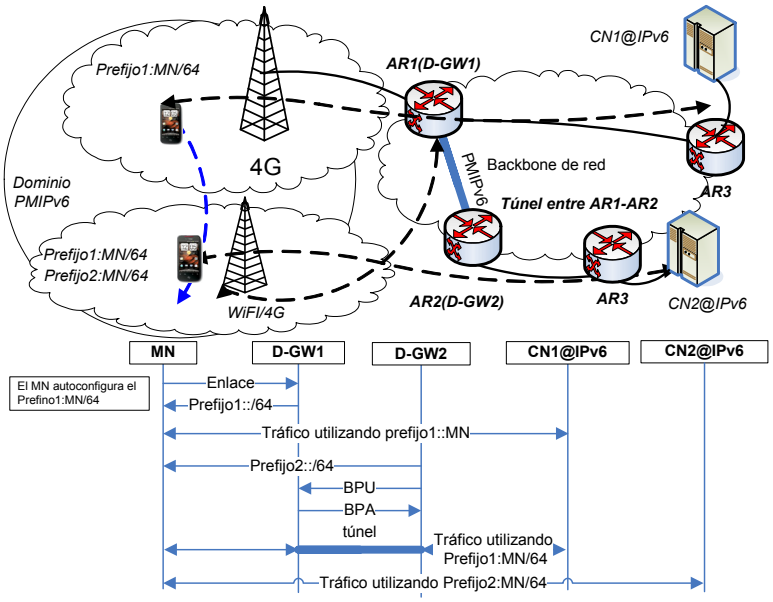

Figura 14. Solución PMIPv6 basada en DMM

El identificador Prefijo1:MN/64 no cambia durante la sesión en curso. El router de acceso AR2 actúa de MAG/LMA para la nueva sesión con el CN2, y de MAG para la sesión con el CN1, mientras que AR1 actúa de LMA para la sesión con CN1.

Los mayores desafíos en el diseño de soluciones DMM se refieren a como un nodo móvil puede recibir y enviar tráfico de forma simultánea entre diferentes gateways D-GW, y como sabe el MN que el prefijo IPv6 asignado esta ocupado, esto es que pertenezca ya a una sesión en curso. Para ello, se introduce una nueva interfaz en el D-GW denominada *Distributed Logical Interface* (DLIF). Cada D-GW anuncia a un MN diferentes prefijos de red. Uno por cada D-GW alcanzable por ese D-GW. Visto así, el D-GW tiene dos roles, en primer lugar actúa como un router, y en segundo lugar actúa como un punto de anclaje en la red.

La figura 15 ilustra el movimiento de un MN1 que se une con el gateway D-GW2 [27]. El D-GW2 anuncia el PrefijoB::/64 hacía el MN1. El MN1 mantiene viva la sesión a través de ese interfaz, interfaz por donde se encamina todo el tráfico hacia el MN1. El D-GW2 a su vez encamina hacía ese interfaz todos los paquetes del interfaz virtual D-GW1-MN1 procedentes de D-GW1. Lo bueno de la idea es que el MN1 solamente tiene un prefijo de red válido en un instante de tiempo determinado.

De ese modo, es posible dar respuesta a necesidades basadas en DMM-PMIPv6, esto es que el MN pueda recibir tráfico de diferentes AR, y saber que router de acceso está suministrando la información sin necesidad de hacer cambios en el software del MN. La solución oculta al MN de tareas de señalización y deja a la red esa responsabilidad.

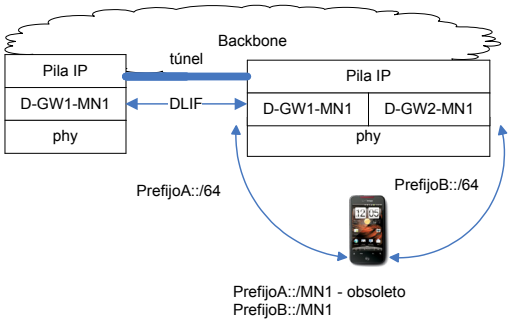

Figura 15. Mecanismo DLIF para anuncios de ruta

Modelo basado en el encaminamiento dinámico. El procedimiento tiene un enfoque diferente. En este caso, cuando un MN se attachea a un AR obtiene una dirección IPv6 que se anuncia utilizando el protocolo Internal *Border Gateway Protocol* (iBGP). Cuando el MN cambia de AR el AR actualiza la información de encaminamiento para alcanzar la nueva ubicación del MN. El método se enfrenta a la convergencia de rutas lo cual provoca enorme latencias, por el contrario la solución propuesta consigue una ruta óptima. De ese modo, la disponibilidad del MN esta asegurada. La solución DMM presenta inconvenientes, como las altas latencias en el proceso de handover debido a la convergencia de las tablas de encaminamiento, también posee además problemas de escalabilidad.

Los mensajes de actualización de ruta pueden llegar a ser excesivos, circunstancia que no es adecuada para conseguir la convergencia del protocolo, y latencia en el handover.

3.2.3. Conclusiones del uso de protocolos IETF en arquitecturas DMM

La señalización y los túneles ocasionan overhead, aunque el mecanismo no siempre es malo. Un handover dentro de un esquema DMM implica crear un nuevo túnel para todas las sesiones en curso. Encapsular significa añadir los 40 bytes de la cabecera IPv6 necesarios para crear un túnel IPv6-to-IPv6. Tal como se argumentó en apartados anteriores en arquitecturas DMM todas las nuevas sesiones están exentas de encapsulamieno de tráfico [28] [29].

Desde el punto de vista de una solución basada en host utilizando el protocolo MIPv6, el encapsulamiento es más severo y dañino. El encapsulamiento de la cabecera túnel viaja por el canal Wireless, muy posiblemente el MN este ubicado en una red de acceso.

En cambio, las soluciones DMM basadas en red con PMIPv6 son capaces de soportar video en tiempo real, videoconferencias, y juegos on-line como consecuencia de las bajas latencias que se producen durante el traspaso. Las soluciones MIPv6 basadas en DMM soportan video en tiempo real siempre y cuando sea tolerante al retardo: como VoD con buffers grandes. Las soluciones tienen un buen rendimiento en términos de latencia, y pérdida de paquetes durante el handover.

Por ultimo, las soluciones basadas en el encaminamiento dinámico son más eficientes, aunque no válidas para soportar tráfico en tiempo real, tienen un rendimiento superior desde el punto de vista de encaminamiento de paquetes, aunque causan excesiva señalización debido a la propagación de rutas. La solución no se puede utilizar en entornos móviles.

3.3. Modelo DMM en redes 3GPP

La mayoría del tráfico fluye mayoritariamente por las redes de acceso, en algunos casos no es necesario encaminar parte de ese tráfico hacia el LMA, posiblemente el tráfico se curse en la red de acceso, en otros casos parte del tráfico no requiere de movilidad IP. De ese modo, es suficiente descargar el tráfico directamente por la pasarela de la red de acceso sin encaminarse hacia el core de red. Desde ese punto de vista es posible encontrar nuevas soluciones para resolver el problema del retraso durante el movimiento del MN en las redes de acceso. Por ejemplo, el LMA puede ser quien decida que flujos/streams IP son encaminados hacía la red origen, y que flujos IP pueden descargarse directamente desde la red de acceso.

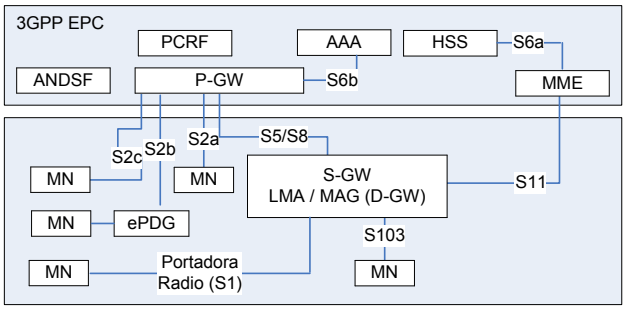

Figura 16. Movilidad interEPC

Desde el punto de vista de la arquitectura EPC, las entidades que deben analizarse en un despliegue DMM son el S-GW y el P-GW. El S-GW es el interfaz hacia el RAN, esto significa que es el punto de anclaje móvil del MN, lugar donde el identificador de red cambia como consecuencia del movimiento del MN en las estaciones base. La figura 16, representa los interfaces analizados en un esquema de movilidad inter-EPC. El interfaz S2a representa accesos non-3GPP trusted, y el interfaz S2b representa accesos non-3GPP untrusted. El interfaz S2c opera bajo DSMIPv6. En todos los interfaces, para resolver el problema se aplican métodos DMM, se consigue que los puntos de anclaje puedan ubicarse cerca del MN, en algunos casos se unifica en una sola entidad el LMA y el MAG.

Las especificaciones propuestas por el 3GPP siempre tratan de resolver el problema de la latencia minimizando el valor, para ello contienen mecanismos conseguir reducir el retardo durante el movimiento del MN sin cambiar el identificador de red, aunque el excesivo retardo que se consigue sigue siendo el punto de inflexión. De ese modo, las arquitecturas evolucionan hacía esquemas multi-acceso dinámicos. De tal modo, que los gateways se colocan en las redes de acceso. El gateway D-GW en la red de acceso se comporta como un AR y adopta elementos de señalización de red para crear los túneles MIP/PMIP. Esto es en el caso de PMIP, operan como un MAG para distribuir el tráfico hacia el MN, y como un LMA para señalizar los túneles PMIPv6. El D-GW se coloca al final del interfaz S5. Una solución parecida se adopta en el modelo DMM, ubicar el D-GW en el interfaz S2a, S2b, y S2c, para accesos 3GPP, untrusted y trusted.

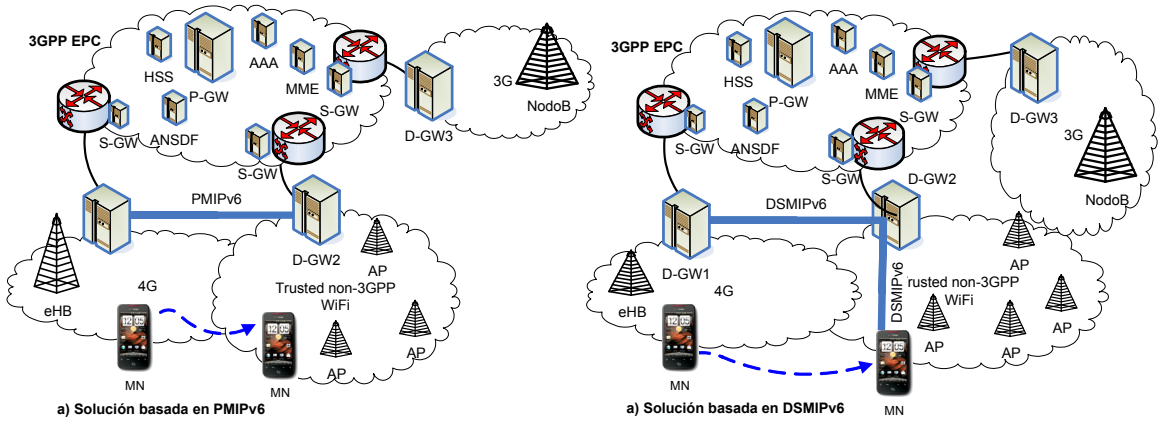

Figura 17. Propuesta DMM que optimiza latencias en un handover

La figura 17 ilustra la arquitectura propuesta para soluciones PMIPv6 y DSMIPv6. La solución optimiza la latencia y los paquetes perdidos. La propuesta reubica el LMA/MAG dentro de las redes de acceso. La mejora en términos de latencia es notable. El D-GW de la llegada del MN. El D-GW no manejará en ese caso demasiados MNs. El D-GW puede asegurar la conectividad con el MN mediante el interfaz virtual. Las conexiones *Packet Data Network* (PDN) se crean y mantienen desde el P-GW siendo transparentes para el D-GW. Para accesos a través del interfaz 2Sb el D-GW puede ubicarse en el ePDG.

La figura 18 ilustra el procedimiento cuando un MN se une a un D-GW y posteriormente realiza un handover. Cuando el MN solicita un acceso se crea un contexto PDN, en ese momento el gateway distribuido asigna un prefijo IPv6 para la nueva sesión, el MN autoconfigura su dirección IPv6. Cuando el MN envía tráfico al gateway, el tráfico no se encamina hacia el LMA, debido a que el D-GW actúa como si fuera un LMA. La figura ilustra el enlace de un MN al D-GW1, el procedimiento disminuye significativamente el overhead.

En un instante determinado el MN realiza un movimiento y alcanza el gateway D-GW2, la sesión PDN pasa a ser gestionada por el D-GW2 actuando como MAG. De ese modo, el MAG negocia un túnel PMIPv6 con el AR anterior que desempeña el rol de LMA en PIMPv6. Circunstancia que permite al MN seguir utilizando el mismo prefijo de red en la nueva ubicación, aunque en realidad para el MN todo el proceso es transparente, debido a que el prefijo en la nueva ubicación quedará obsoleto para el MN. El MN autoconfigura entonces una nueva dirección IPv6 a partir del prefijo de red que le suministra el nuevo MAG (D-GW2).

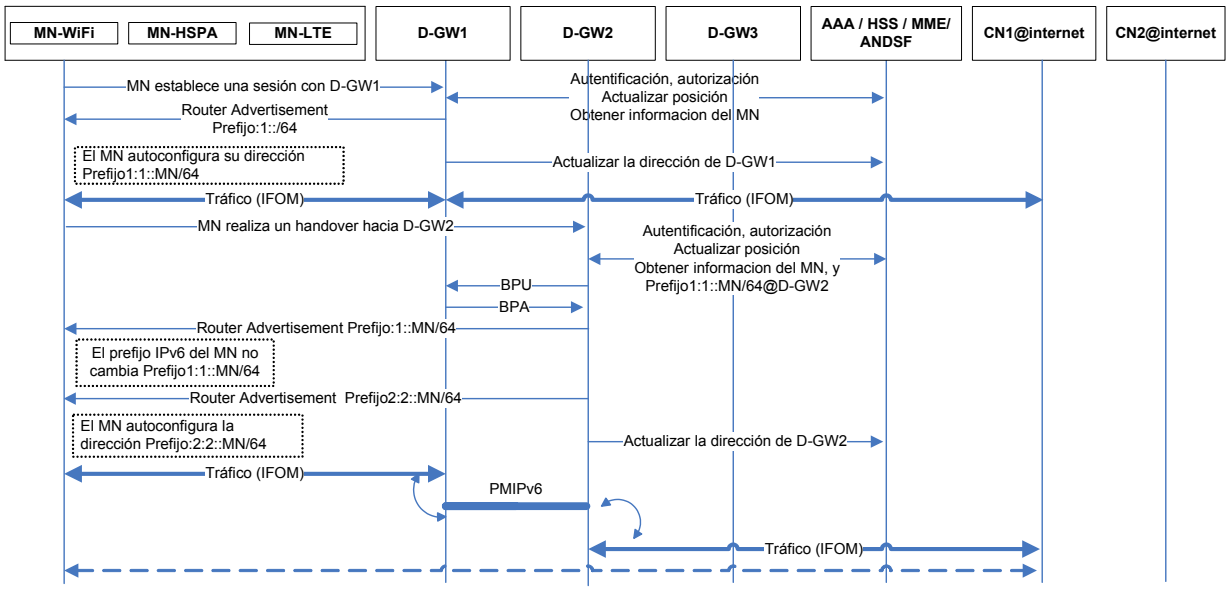

Figura 18. Procedimiento para señalizar un handover PMIPv6 en una arquitectura DMM

4. Solución transparente de movilidad con optimización de QoS

Las arquitecturas 6LSA surgen como consecuencia de la continúa búsqueda de nuevos métodos para optimizar la infraestructura utilizando tecnologías de conmutación de paquetes. Arquitecturas que no se complementa con MPLS. Por otro lado, las redes de conmutación de paquetes no se diseñaron con la mente puesta en su gestión, más bien fueron infraestructuras donde desplegar servicios, y experimentar protocolos que adaptan automáticamente el tráfico a la infraestructura. Nadie buscó en sus inicios mecanismos capaces de minimizar la congestión, o reducir latencias para servicios sensibles al retardo, en un principio no se pensó en adaptar el tráfico a las condiciones de la infraestructura, debido al gran coste en señalización que conlleva. La señalización aumenta la complejidad en la gestión del sistema. Por este motivo, las nuevas propuestas pasan por dotar de semántica a los protocolos de capa 3.

La gestión de las políticas de red es altamente dinámica, debido a que se añaden nuevos dispositivos continuamente, los requisitos de servicio cambian rápidamente. Tal circunstancia obliga a buscar nuevos mecanismos que faciliten la gestión de la infraestructura de red.

Surgen entonces iniciativas que trasladan el plano de señalización hacia data centers virtualizados, capaces de separar las reglas del servicio de la arquitectura subyacente. De ese modo, es posible efectuar una gestión de recursos teniendo una visión global de todo el sistema, es posible buscar y hacer algoritmos eficientes que junto a técnicas big data permiten obtener información predictiva con la que aprovisionar los recursos y conseguir un trato más equitativo de los mismos en términos de QoE. Entre las decisiones que pueden tomarse están el coste en la red de acceso, la posibilidad de identificar los mejores accesos disponibles, la QoS en la capa de enlace, frecuencias utilizadas, tasa de bit máxima permitida, operadores existentes, el impacto del handover en el servicio cursado.

En este sentido, emergen *Software Defined Networking* (SDN), la idea es ver la red como un pool de recursos que son asignados en función de las necesidades del servicio. Cada nodo de red expone un interfaz común independientemente del fabricante. La idea simplifica mucho, la gestión de los nuevos servicios de banda ancha, mediante la caracterización de los flujos que recorren una ruta a partir de un perfil de servicio, tal circunstancia permite aislar e independizar servicios, mejora especialmente los sistemas autónomos. Esto es debido a que los sistemas autónomos pueden pertenecer a distintos tipos de fabricantes, tanto hardware como software. Actualmente, aplicar QoS, o transportar un determinado flujo E2E en sistemas legados, no es trivial ni fácil. Por ese motivo aparecen nuevos métodos capaces de simplificar o abstraer la complejidad a través de un interfaz común e independiente del fabricante y tecnología, es posible sustituir nodos muy específicos y propietarios del fabricante por nodos genéricos con interfaces estándares. La idea consiste en disminuir la complejidad en la gestión del sistema para trasladar fácilmente políticas del usuario a parámetros de red [30].

El enfoque actual para proveer QoS en flujos de tráfico está basado en el carácter rígido que siguen los datagramas durante su ruta, esto significa que la red es quien establece una ruta para un determinado datagrama. En este sentido, no existe posibilidad de añadir una capa superior que determine o modifique las reglas que debe seguir un paquete durante su ruta.

En esquemas SDN solamente se señalizan los paquetes que los nodos no saben como conmutarlos. La señalización en esquemas SDN se realiza a través de controladores centrales que gestionan el aprovisionamiento de recursos en cada nodo a través del interfaz que expone el nodo y utilizando el protocolo OpenFlow. El controlador señaliza el nodo switch por medio del protocolo inyectando información en el nodo switch.

Antes de definir la solución, analizamos brevemente los mecanismos de gestión de recursos actuales. Actualmente dos son los retos que deben superar los enlaces, en primer lugar los problemas ocasionados por la congestión, suponiendo que la congestión no existiera, sería suficiente sobredimensionar los enlaces mecanismo poco eficiente en realidad, de ese modo

debemos asumir la congestión de red como algo natural y gestionarla. El segundo reto se basa en mantener la calidad de servicio antes y después del handover entre diferentes dominios administrados, el cambio no debe significar una degradación de servicio. Por lo tanto, los mecanismos de gestión de la movilidad deben ser capaces de proporcionar los recursos que el usuario espera de la red con el fin de aprovisionar eficientemente el servicio.

Surgen entonces mecanismos de aprovisionamiento de servicios como *Integrated Services* (IntServ/RSVP) que reserva recursos a priori en la infraestructura que proporciona el servicio. *Differentiated Service* (DiffServ) que diferencia los diferentes tipos de servicio en la infraestructura y la conmutación de etiquetas MPLS capaz de segregar el tráfico mediante etiquetas. MPLS separa la parte de encaminamiento de la parte de conmutación en el reenvío de los paquetes. La parte de encaminamiento es compleja y lenta en cuanto a tiempos de convergencia y cálculo de rutas. La parte de conmutación es rápida y simple. Se realiza a partir de información existente en la *Forwarding Information Base* (FIB) de cada nodo. El planteamiento y funcionamiento encaja con los objetivos de diseño de OpenFlow tecnología capaz de conmutar paquetes a partir de reglas predefinidas, y capaz de aislar flujos similar a tecnologías como IEEE 802.1Q.

Por otro lado, se denomina Ingeniería de Tráfico (TE) a la planificación de rutas en una red a partir de previsiones y estimaciones a largo plazo con el fin de optimizar los recursos y reducir la congestión, no es más que decidir cómo se encamina el *Label Switched Path* (LSP) a través de un dominio MPLS. La tecnología MPLS reduce de forma significativa el procesamiento de los paquetes en los routers, permite diferenciar servicios a través de una etiqueta de cabecera. El encaminamiento del tráfico se realiza mediante la etiqueta de cabecera. Actualmente es la tecnología que resuelve el problema del encaminamiento con mayor eficacia. Para ello crea rutas optimizadas y busca la forma de organizar ese tráfico para que no fluya por rutas poco optimizadas. MPLS permite en cierto modo que el flujo siga una única ruta creando, para ello un LSP cuando sea necesario, del mismo modo se puede restringir el throughput cuando sea necesario o cambiar el LSP en caso de movimiento del MN [31].

La propuesta que proponemos en el trabajo se basa en integrar MPLS junto a SDN. Por otro lado, las redes actuales están lo suficientemente maduras como para introducir el concepto de SDN.

En SDN cuando un nodo recibe un paquete comprueba la etiqueta de cabecera que identifica el flujo, en el caso de que el valor sea cero el proceso de encaminamiento se realiza normalmente, en cualquier otro caso en nodo busca en la *Ternary Content Addressable Memory* (TCAM) información para conmutar el paquete. La información contenida en la tabla TCAM esta temporizada. Si el paquete no tiene una regla en la TCAM el nodo preguntará al controlador por la tupla que identifica el paquete, y las acciones asociadas. El nodo envía eventos periódicamente al controlador para actualizar su estado.

La información en la TCAM esta agrupada en reglas. Las reglas están agrupadas y juntas forman una tupla. (Ver tabla 2). La interfaz consigue abstraer el plano de control. Cuando llega un paquete a un nodo se realiza una comprobación con algún campo de la tupla, si coincide la tupla que identifica al paquete de llegada, esta contenida en la TCAM se desencadena una acción. Si coinciden varias reglas se utiliza el campo prioridad. Las nuevas versiones de OpenFlow 1.2 soportan *matching* de direcciones fuente y destino IPv6, soportan la etiqueta de flujo IPv6, el campo clase de tráfico, ICMPv6ype, ICMPv6.

Cabecera											**Acciones**
Puerto entrada	Prioridad	VLAN ID	Ethernet			IPv4			Transporte		Borrar, modificar, redireccionar
			sa	Da	type	sa	da	protocolo	src	dst	

Tabla 2. Tupla definida en OpenFlow 1.0

La arquitectura SDN esta formada por tres componentes, los nodos OpenFlow switch, el protocolo OpenFlow y el controlador. Los nodos OpenFlow switch son unidades software instaladas en los diferentes nodos de red, contienen una tabla que almacena una acción por cada flujo, las acciones las proporciona el controlador a través del protocolo OpenFlow. Posibles acciones son borrar el paquete, modificar la cabecera del paquete, como la dirección destino, conmutar el paquete. El protocolo OpenFlow señaliza cada uno de los nodos de red. Todos los nodos se conectan al controlador para señalizar el tráfico utilizando canales seguros con IPSec. La especificación define mecanismos tanto asíncronos como síncronos.

Los controladores están compuestos por módulos llamados *Network Operating System* (NOX), cada módulo lleva a cabo funciones diferentes según el tráfico, el mecanismo permite escalar la arquitectura, existe por ejemplo un modulo de gestión de recursos radio, gestión de movilidad, firewall, encaminamiento, MPLS, etc. De ese modo, cada controlador puede especializase en un tipo de tráfico.

La estructura SDN permite abstraer entidades del plano de control y moverlas hacía un controlador virtualizado. El controlador a partir de la información que le proporcionan los nodos e información almacenada y aprendida suministra a los nodos información por donde señalizar las rutas para cada flujo. Entre las decisiones que el controlador puede tomar están en primer lugar los parámetros de codificación del flujo, el codec utilizado, el frame rate de compresión, finalmente suministrar los requerimientos del enlace por ejemplo la tasa de bit que demanda el servicio. El modelo propuesto asegura una menor latencia, optimiza los enlaces de acuerdo a las exigencias del servicio, debido a que las decisiones se toman teniendo en cuenta la capa de enlace, la capa de red y la capa de transporte.

Actualmente OpenFlow no implementa capacidades MPLS. En un dominio MPLS el tráfico se clasifica según criterios de calidad de servicio, los encaminadores no realizan ningún otro análisis que procesar la etiqueta de cabecera a partir de la tabla FIB. Todo el reenvío de tráfico se basa en etiquetas, los switches solamente comprueban la etiqueta y no la dirección destino, el mecanismo permite mejorar el rendimiento respecto al reenvío del paquete por procedimientos de capa 3 convencional. Así pues, la redirección MPLS puede ser efectuada por conmutadores capaces de verificar y de reemplazar las etiquetas a una velocidad y una QoS adecuadas. En MPLS no es necesario analizar los encabezamientos de capa IP, circunstancia que produce una ventaja, aunque tiene algún inconveniente como la reserva de recursos salto a salto. La señalización de la ruta a través de OpenFlow mejora tal circunstancia.

MPLS puede soportar dominios o niveles, tal circunstancia permite definir más de un circuito virtual para un mismo paquete. Para ello, MPLS utiliza una pila de etiquetas que van encapsuladas en la cabecera del paquete. Las decisiones de routing siempre se realizan a partir de la etiqueta contenida en la cima, última etiqueta de la pila primera que se procesa. De ese modo, es posible anidar etiquetas para crear por ejemplo VPNs. Los paquetes viajan según los valores de esas etiquetas. El LSP son rutas que se establecen dentro de una red MPLS, los LSPs se forman desde el destino hacia el origen. El origen inicia una cadena de mensajes de petición de etiquetas para crear un LSP. El destino responde con mensajes de asociación de etiquetas creando el LSP. De ese modo se va formando el LSP hasta el origen. Según las especificaciones del IETF, MPLS debe funcionar sobre cualquier tipo de transporte: *Point to Point Protocol* (PPP), *Local Area Network* (LAN), ATM, FR u otros.

Actualmente en dominios MPLS la reserva de recursos se lleva a cabo a través del protocolo RSVP-TE (RFC 4804), ó *Label Distribution Protocol* (LDP). El protocolo señaliza los túneles LSP, para ello cada nodo rellena las tablas FIB. El protocolo permite reenrutamiemto de túneles LSP ante una caída de la red, congestión o un cuello de botella. El funcionamiento del protocolo es como sigue: el router de entrada determina las necesidades del túnel LSP con el router de salida. En este punto se establecen las políticas de la conexión, y se fija quien será el router de salida. El nodo de entrada prepara y envía un mensaje de tipo PATH. Los nodos

intermedios reciben el mensaje de tipo PATH, comprueban que no son el router de salida y reenvían el mensaje hacia el próximo salto. El procedimiento se repite hasta que el mensaje alcanza el router de salida. Cuando el nodo de salida verifica que es el aprovisiona los recursos solicitados, elige una etiqueta y interfaz de salida y guarda esa información en la FIB.

La etiqueta se distribuye dentro de un mensaje de tipo RESV en el campo LABEL. El mensaje se envía por el puerto que llego el mensaje PATH. Los nodos intermedios reciben el mensaje de tipo RESV, y guardan ese puerto de salida junto a la etiqueta en la tabla FIB, a continuación preparan un mensaje para el nodo anterior, eligen un puerto de salida y una etiqueta que también guarda en la FIB. Cuando el mensaje llega al nodo origen, el nodo solamente actualiza la tabla FIB no añade ninguna etiqueta.

Por ejemplo, elige la etiqueta 99 para el tráfico dirigido a la IP 102.102.102.102. Posteriormente envía un mensaje al router anterior. El router anterior recibe una etiqueta de salida 99 y elige una etiqueta 76 como etiqueta de entrada. Posteriormente envía ese valor al router anterior, así sucesivamente hasta que se alcanza el router de entrada. El router de entrada recibe la etiqueta que pidió, y de forma implícita se crea la ruta LSP.

La conmutación de etiquetas en un LSR se lleva a cabo del siguiente modo. En primer lugar se examina la etiqueta del paquete entrante y la interfaz por donde llega, posteriormente se consulta la tabla de etiquetas FIB para saber la interfaz de salida y la etiqueta se debe añadir. El paquete se encamina hacia el próximo LSR.

4.1. Mejora de tráfico con OpenFlow y MPLS en el EPS

En un dominio MPLS es posible que el MN cambie de router de acceso (*Mobile Access Router* (MAR)). Por lo tanto el LSP que estuviera establecido debe liberarse, y crearse un nuevo LSP hasta el nuevo MAR. El hecho de liberar un LSP y crear otro nuevo en cada movimiento resulta muy costoso en términos QoS, y no parece ser la mejor solución cuando se persigue minimizar el tiempo para restablecer la conectividad del servicio durante el handover.

Por esta razón, proponemos que el túnel LSP tenga una parte fija y otra móvil. La parte móvil cambia en cada movimiento del MN en el router de acceso. La provisión de QoS solamente se negocia en el nuevo LSP. Por otra parte, al utilizar mecanismos de ingeniería de tráfico sobre una red MPLS dentro de esquemas SDN, es posible aplicar distintas restricciones durante el cálculo de las rutas totales o parciales. En un dominio MPLS la ruta se crea dinámicamente según se describió en párrafos anteriores. La tecnología MPLS introduce una etiqueta que representa al *Forwarding Equivalence Class* (FEC), conjunto de paquetes que entran en un dominio MPLS por la misma interfaz reciben las mismas etiquetas. La cabecera MPLS tiene un tamaño de 32 bits. Los 20 últimos bits están reservados para albergar la etiqueta, los primeros 12 bits se comportan para mecanismos QoS, TTL o la existencia de anidamiento de cabeceras, que se indica con el bit S. (Ver figura 19).

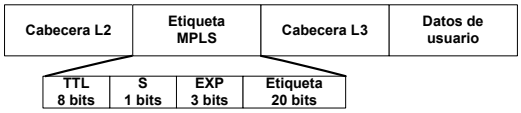

Figura 19. Etiqueta MPLS

Para conmutar los paquetes en OpenFlow se consulta la tabla TCAM. Si la ruta no está aprendida en el nodo switch se pregunta al controlador. Por lo tanto, el concepto de flujo en MPLS no es tan genérico como en SDN en términos de definición de flujo y reglas asociadas. En SDN un flujo es una asociación entre paquetes que son parte de un stream de comunicación, todos los paquetes reciben el mismo tratamiento por la red. En MPLS el concepto de flujo se refiere a una FEC que crea una ruta LSP, por donde se enviarán los paquetes que caracterizan el flujo dentro del dominio administrado. En SDN un nodo switch trata cada paquete de modo independiente del resto de nodos, por el contrario en MPLS primero se debe negociar

previamente un LSP y este se mantendrá fijo durante toda la sesión, el LSP es la ruta donde se enviarán los paquetes.

En definitiva en MPLS-TE cada nodo debe decidir antes de encaminar el flujo, si es posible a priori reservar recursos a partir de la información proporcionada por los protocolos de enrutamiento, en cambio en SDN la reserva se lleva a cabo a través del controlador, quien tiene información de todo el sistema. La solución consigue disminuir el número de ciclos de cpu en los nodos como consecuencia de la ausencia de información de los protocolos de enrutamiento.

La propuesta del trabajo estudia como extender la *flow table* en cada nodo switch para que contenga las etiquetas MPLS. La información sobre las etiquetas de cabecera las inyecta el controlador. El tamaño de la etiqueta en MPLS no tiene un límite, esto significa que se pueden anidar varias etiquetas MPLS, en el ejemplo se establecen dos pero pueden anidarse más. El bit S puesto a uno indica que es la cabecera que está más arriba, en el resto de cabeceras el bit S vale 0. El enfoque permite transportar varias cabeceras MPLS encapsuladas, una para identificar por ejemplo el servicio VPN y la otra para identificar el túnel de transporte. Un túnel VPN es diferente a un túnel de transporte, por ejemplo este último no requiere de autentificación.

En el plano de control cada nodo implementa tres acciones. Push, colocar o insertar si no existe una etiqueta en la cima de la pila de etiquetas. Pop, quitar la ultima etiqueta MPLS de la pila. Swap, intercambiar etiquetas MPLS.

Puerto Entrada	Prioridad	VLAN ID	Ethernet			MPLS		IPv4			Transporte	
			sa	da	type	label 1	label 2	sa	da	protocolo	src	dst

Tabla 3. Extensión en la tupla OpenFlow para contener las etiquetas MPLS

MPLS OpenFlow tiene tres tipos de nodos ingress (push), transit (swap), egress (pop). Las acciones MPLS asociadas son: push_mpls, pop_mpls, sawp_mpls, decrement_ttl, copy_bits. Push_mpls, coloca la etiqueta MPLS de 32 bits en la cima de la etiquetas MPLS del paquete, también copia el campo TTL y el campo QoS desde la cabecera IP o cabecera MPLS anterior a la cabecera MPLS actual. Pop_mpls, quita la cabecera MPLS de la cima del anidamiento de etiquetas MPLS, y copia los bits TTL y QoS hacia la cabecera IP o etiqueta MPLS que tuviera el paquete debajo de la actual. Swap_mpls, intercambia los 20 bits entre las dos cabeceras MPLS superiores. Decrement_ttl, decrementa el valor del campo TTL, borrando el paquete si hubiera expirado. Copy_bits, copia los bits del campo TTL y Qos desde/hacia la cabecera IP o hacia la siguiente etiqueta MPLS.

4.1.1. Ejemplo de una arquitectura IP/MPLS a través de esquemas SDN

Cada servicio requiere caracterizar el tráfico de un modo diferente, por ejemplo por dirección fuente y destino, en otros casos monitorizar el tráfico por la dirección fuente, etc. Otros servicios, requieren reglas de control de admisión, eliminar cierto tráfico, cambiar la dirección de origen, o balancear el tráfico a través de diferentes rutas. Para ello OpenFow introduce mecanismos capaces de programar estos comportamientos a través de sofisticadas aplicaciones capaces de gestionar el tráfico en real-time. El objetivo permite dotar de semántica y comportamiento al protocolo, para permitir adaptar los flujos de tráfico a los recursos físicos de infraestructura. La idea es equilibrar de forma óptima la utilización de los recursos de la infraestructura.

En MPLS el LSP se utiliza para trazar la ruta desde un nodo de entrada hasta el nodo de salida. Los LSRs forman el LSP a través de señalización en el plano de control, cada LSR conmuta los paquetes a partir de la información contenida en la *flow table*. En esquemas SDN no se utilizan protocolos de reserva de recursos como Label Distribution Protocol (LDP). En el plano de datos solamente se utiliza la tecnología MPLS para transportar datos, en el plano de usuario se utiliza OpenFlow.

Cuando un MN establece una sesión con el CN, envía en la petición un paquete marcado con un valor FID. El valor lo asigna el MAR durante la creación del contexto. La marca viaja en el campo de cabecera IPv6. El valor FID es único, para ello se realiza un hash con los campos de cabecera etiqueta de flujo, dirección fuente y dirección destino. En controlador utiliza el valor del FID para crear un contexto para ese flujo. Cada nodo define y mantiene un contexto con el controlador a través del identificador de flujo. El controlador también mantiene el contexto. Contexto que sincroniza síncronamente, ó asíncronamente con cada nodo. El contexto se asocia a cada paquete de modo automático mediante el identificador FID. El mecanismo obtiene resultados excelentes para servicios en tiempo real. El nodo intercambia información de monitorización con el controlador a través de eventos, e informa al controlador del estado del flujo. Si las características QoS del flujo estuvieran por debajo de un determinado umbral, se envían alarmas al controlador quien señaliza nuevas rutas. Los handovers son predictivos a partir de información histórica de usuario.

Puerto de entrada	Etiqueta de entrada	FID	Puerto de salida	Etiqueta de salida	Timeout	QoS
3	65	453	5	55	1567	BE
2	73	523	5	55	3453	PRI
1	82	987	4	35	1134	NOP

Tabla 4 Contenido de la tabla TCAM en el R3

En redes SDN de alto rendimiento, las decisiones de conmutación las toman los nodos switch MPLS en base a la *flow table*, son mucho más sencillas y rápidas que las que toma un router IP ordinario. La anidación de etiquetas permite agregar flujos con mucha facilidad, por lo que el mecanismo es escalable.

Cuando el MN realiza un handover, el nuevo router de acceso señaliza el procedimiento con el controlador, reallocando el LSP en el nuevo router de acceso. Los MARs detectan el MN por medio de procedimientos de descubrimiento de ruta de capa 3 o triggers en capa 2. La figura 20 ilustra el procedimiento para reallocar el túnel durante el movimiento del MN.

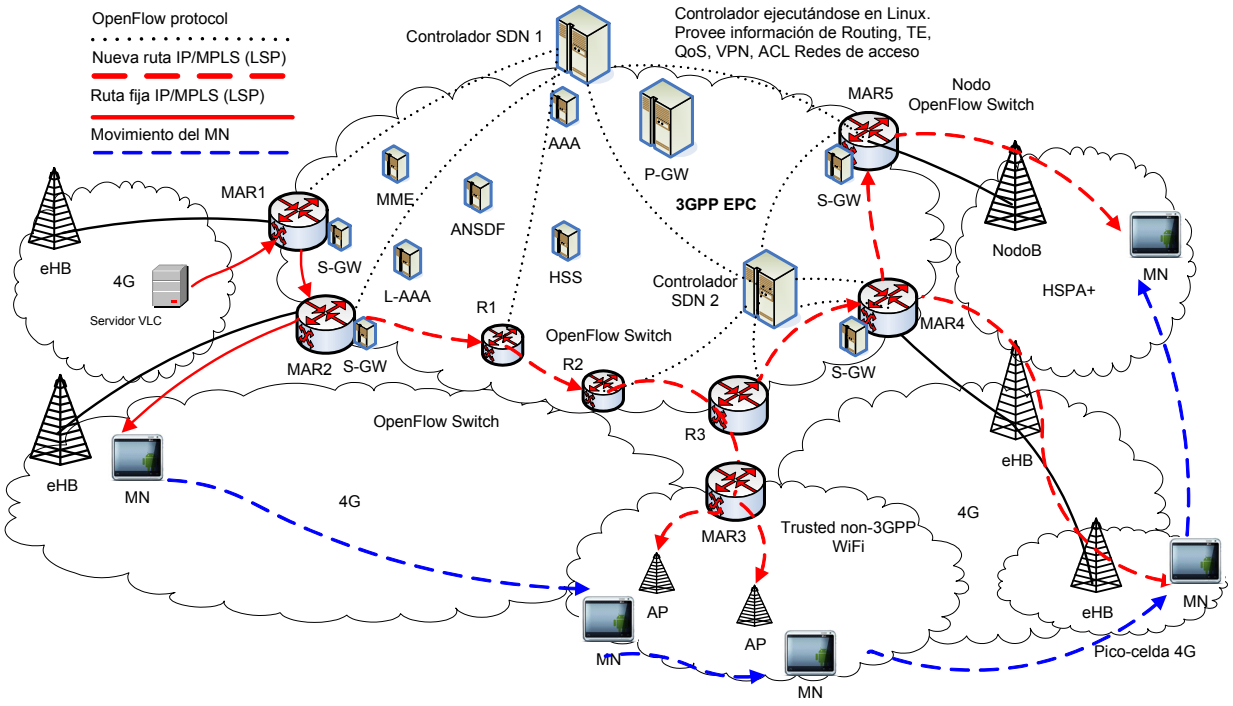

Figura 20. Prototipado de una arquitectura basada en IP/MPLS bajo esquemas SDN.

4.2. Entorno de pruebas para evaluar la arquitectura de movilidad

En este apartado modelamos la arquitectura de la Figura 20 mediante un entorno de pruebas donde experimentar, y poder disponer de un escenario de pruebas sobre el que poder realizar

diferentes tipos de test. Para elegir el escenario se han seguido los siguientes pasos con el fin de disponer un entorno para emular la arquitectura de red. En primer lugar, se contemplo la posibilidad de coexistir varias redes en un mismo entorno, en segundo lugar la convergencia entre tecnologías, y finalmente la posibilidad de conectividad global entre todos los elementos que forman la red [32].

La emulación de los nodos se realiza con software open source mininet-2.0.0-113012-amd64-ovf.zip, software capaz de emular topologías de red de modo sencillo [32]. El software soporta el estándar OpenFlow, esto significa que tiene interfaces amigables para programar los nodos switches. El software permite emular una arquitectura de red a través de esquemas SDN en un único PC. El procedimiento es útil para experimentar nuevas arquitecturas de red. Mininet incluye interfaces python para interactuar con la arquitectura dentro de una máquina virtual en virtualbox 64amd y Ubuntu 12.04 LTS. Mininet permite parametrizar el ancho de banda del enlace, el porcentaje de paquetes perdidos, el tamaño del buffer y la latencia. Mininet incorpora *ipref* para estimar anchos de banda entre enlaces. Las reglas se inyectan a través del controlador POX [33]. POX es un controlador escrito en python que permite implementar de forma fácil comportamientos en los nodos de la topología de red, permite también enlazar los interfaces fisicos y virtuales mediante xterm [34] [35].

El ancho de banda que se configura en los enlaces 4G es de 15Mbps, con latencia es de 10ms. La configuración del ancho de banda en enlaces 3G es de 2Mbps, con latencia de 50ms. Los enlaces WiFi tienen un ancho de banda de 54Mbps y una latencia de 15ms. El tamaño de MTU es de 768 bytes. La generación de tráfico se realiza mediante el servidor vlc [45] servidor que actúa con el rol de CN, capaz de generar tráfico TCP videostreaming, por ser tráfico mayoritario en redes actuales. Actualmente hay dos formas de enviar video a través del protocolo http. En primer lugar bajo el modelo *download progresivo*, y en segundo lugar utilizando el modelo *stream adaptativo*. En el test utilizamos un enfoque *Adaptive Scalable Video* (H.264/SVC) *Streaming* sobre HTTP, donde la tasa de bit es capaz de ajustarse al ancho de banda disponible [46]. Para ello, los servidores de video almacenan varias copias del video en formatos diferentes, una para cada stream según la tasa de bit. También se dimensionan los buffers para que actúen como almacenes de paquetes. Los buffer absorben el tráfico que la red no puede gestionar en ese momento jugando un papel importante en la regulación del tráfico

4.2.1. Caracterización del retardo E2E

La figura 21 ilustra las latencias que deben tenerse en cuenta en un *handover* en una red de acceso. El 3GPP define latencia como la suma de los tiempos que transcurren desde que el MAR detecta un MN hasta que se produce la continuidad en la sesión en el nuevo router móvil de acceso. El T*echnical Report* 36.839 v11.1.0 especifica tiempos de 50ms para preparar el *handover* y otros 40ms para realizar el *handover* en redes HetNet.

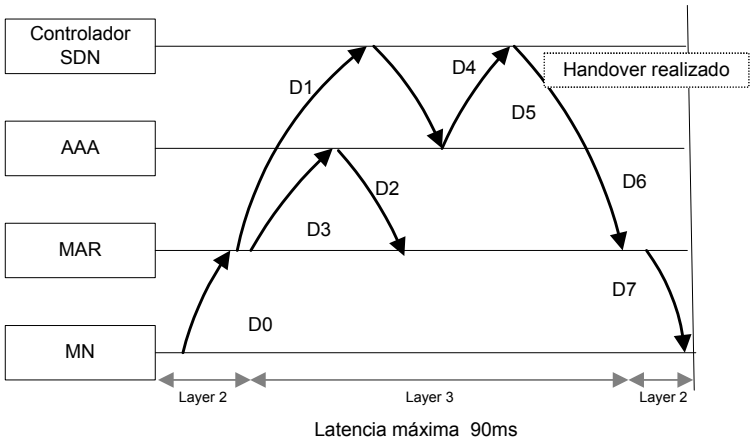

Figura 21. Latencia máxima en un handover (TR 36.839).

Para un servicio VoIP la ITU en su norma G.114 [43] establece ciertos límites para la latencia. Retrasos de mayores de 150ms son muy molestos en una conversación, retrasos mayores de 400 ms son inaceptables. La tabla 5 ilustra la caracterización del retardo E2E para VoIP. [46]

Latencia E2E (ms)	**Calidad**
0-150	Aceptable para todos los usuarios
150-400	Aceptable con gran impacto
> 400	Inaceptable

Tabla 5. Caracterización de la VoIP en términos de retardo. (N.Pavlidou)

Para la generación del tráfico videostreaming se utiliza el formato: CIF (352x288), duración 12sg a 25fps y con 300 frames enviados, la latencia máxima soportada es entorno a 120ms. La figura 22 representa a diferentes frames recibidos con escaso movimiento en el video, se transmiten pocos vectores de movimiento y diferente jitter. Latencias superiores a 100ms con jitter son inaceptables, pero en cambio 116ms de latencia sin jitter es aceptable.

Para mejorar el retardo existen técnicas como priorización de paquetes, mecanismos eficientes en el scheduling, y algoritmos mejorados en los buffers. Hay que tener en cuenta que vlc no soporta reordenamiento de frames, cuando un frame llega fuera de secuencia se borra.

Para el modelo de tráfico utilizamos el modelo *Self-Loading Periodic Streams* (SLoPS). La fuente envía periódicamente información sobre el enlace. Un stream SLoPS se compone de un número de paquetes (K) de longitud (L) a una tasa constante (R). El controlador monitoriza el One-way delay a través de eventos proporcionados por el nodo. De este modo si el bit rate (R) es mayor que el ancho de banda disponible (A) el stream provoca sobrecarga en el enlace aumentando el *One-way delay*. Por el contrario, si el bit rate es menor que el ancho de banda disponible el *One-way delay* permanecerá estable. El método funciona como el clásico algoritmo de búsqueda binaria, la fuente intenta acercar la tasa a la capacidad permitida por el enlace mediante sucesivas iteraciones, iteraciones guiadas a través del *One-way delay*. El mecanismo asegura que el enlace esta trabajando a plena carga.

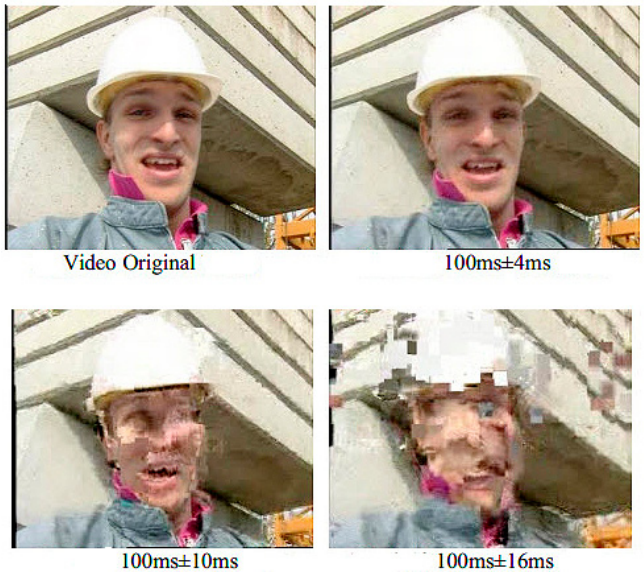

Figura 22. Frames de video recibido para diferentes valores de jitter

El ancho de banda y retardo se miden a partir del método *Variable Packet Size* (VPS) que mide la capacidad del enlace enviando paquetes de diferente tamaño a todos los nodos de la ruta, los nodos responderán con mensajes de control de tipo ICMP, para informar sobre su estado. De ese modo el nodo origen utiliza ese retraso para medir el RTT. Se utiliza el Time-To-Live (TTL) para obligar a los nodos de la ruta a responder. Para un mismo paquete del mismo tamaño se utiliza un RTT más pequeño. El RTT es la suma del tiempo de propagación, el

encolado y la serialización del paquete. El método estima, que uno de cada tres paquetes enviados sufrirá el menor retardo posible.

En primer lugar analizamos la métrica del throughput. El throughput TCP se ve afectado por el número de conexiones simultáneas, por el tráfico UDP que exista en ese momento en el enlace, por el tamaño de los buffers tanto en el emisor como en el receptor, la congestión del enlace, y la capacidad del enlace. Por ejemplo, para una transferencia de una página web donde el volumen de tráfico es muy pequeño, el throughput esta afectado directamente por el tamaño de los buffers, el RTT, y la existencia de mecanismos *slow-start* en el protocolo, antes que el ancho de banda disponible en el enlace. Así pues, el throughput de una transferencia TCP a través de un servicio E2E esta más afectado por la versión del protocolo TCP que por el ancho de banda disponible en el enlace.

Analizamos el retraso de los paquetes debido a que influye significativamente en la calidad del servicio, en algunos casos el servicio es intolerable. El RTT es el tiempo que tarda un paquete en realizar la ruta entre el emisor y receptor. El tiempo se caracteriza por la suma de los retardos de propagación, los retardos de encolamiento, y también los tiempos utilizados por los algoritmos de control de congestión en la capa de transporte. Es importante conseguir medir ese valor, para adaptar las tasas de transmisión al medio.

Finalmente, se analiza los paquetes perdidos. Métrica que determina la calidad del enlace, normalmente los paquetes perdidos no se deben a un tráfico excesivo, si no más bien a algoritmos inadecuados en la cola del buffer, o la existencia de overflow en el buffer. Los paquetes perdidos se estiman a través de la secuencia en la cabecera. (S. Karapanzatis-2009) maneja valores típicos tolerables de 1% hasta el 3% de paquetes descartados dependiendo de la calidad de la transmisión.

4.2.2. Simulación a través de la composición de esquemas SDN

Para demostrar una implementación real y demostrar su viabilidad realizamos un test con tecnología OpenFlow switching capaz de señalizar rutas MPLS. Se pretende demostrar, y analizar la degradación del enlace cuando el MN de la figura 20 cambia de posición en diferentes routers dentro de las redes de acceso. (Ver figura 23).

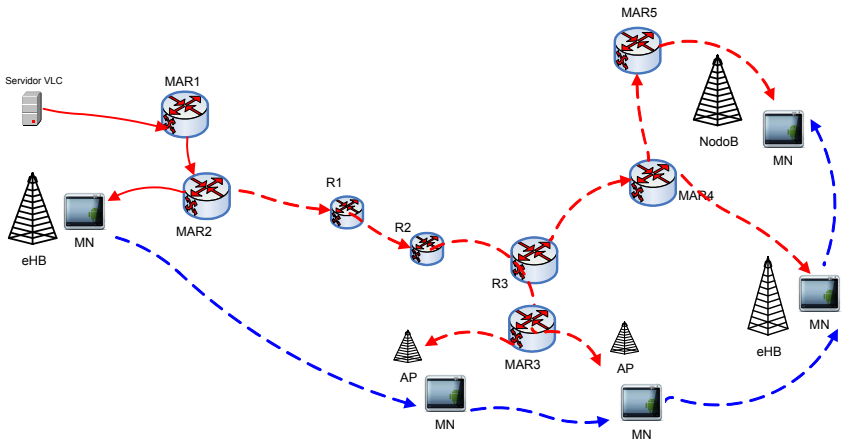

Figura 23. Test de evaluación

OpenFlow permite abstraer las reglas de negocio de la arquitectura, y de ese modo dar visibilidad a las reglas de negocio en el servicio, para ello inyecta las reglas en el nodo switch. Cualquier aplicación software debe estar modularizada, y SDN no es una excepción. En SDN cada modulo se resposabiliza de tareas especificas, por ejemplo un modulo encargado de monitorizar el tráfico, y otro conmutarlo. Cualquier aplicación SDN encargada de gestionar rutas sólo ve una colección de servicios que representan a determinados flujos de tráfico. La encapsulación y desencapsulacion de los paquetes en cada flujo permite presentar los servicios

como paquetes individuales. Los esquemas SDN convergen en muy poco tiempo, de ese modo la perdida de paquetes es nula.

La tabla 6 ilustra diferentes entradas en la *flow table* en el nodo switch R3 después de crear la infraestructura con mininet y ejecutar el algoritmo capaz de aprender rutas en el controlador POX. El POX llena las tablas de los nodos. Cada entrada tiene una etiqueta origen, una etiqueta destino, un puerto de entrada y destino por donde conmutar los paquetes.

cookie=0, duration_sec=1s, duration_nsec=6000000s, flow_id=453, table_id=0, priority=65535, n_packets=1, n_bytes=98,idle_timeout=10,hard_timeout=30,icmp,in_port=3,dl_vlan=0xffff,dl_vlan_pcp=0x00,label_src=65,label_dst=55,nw_src=10.0.0.2,nw_dst=10.0.0.10,nw_tos=0x00,icmp_type=8,icmp_code=0,actions=output:5
cookie=0, duration_sec=0s, duration_nsec=716000000s, flow_id=523, table_id=0, priority=65535, n_packets=1,n_bytes=98,idle_timeout=10,hard_timeout=30,icmp,in_port=2,dl_vlan=0xffff,dl_vlan_pcp=0x00,label_src=73,label_dst=55,nw_src=10.0.0.3,nw_dst=10.0.0.10,nw_tos=0x00,icmp_type=8,icmp_code=0,actions=output:5
cookie=0, duration_sec=0s, duration_nsec=843000000s, flow_id=987,table_id=0, priority=65535, n_packets=1, n_bytes=98,idle_timeout=10,hard_timeout=30,icmp,in_port=1,dl_vlan=0xffff,dl_vlan_pcp=0x00,label_src=82,label_dst=35,nw_src=10.0.0.2,nw_dst=10.0.0.3,nw_tos=0x00,icmp_type=8,icmp_code=0,actions=output:4
cookie=0, duration_sec=0s, duration_nsec=838000000s,flow_id=453,table_id=0, priority=65535, n_packets=1, n_bytes=98,idle_timeout=10,hard_timeout=30,icmp,in_port=5,dl_vlan=0xffff,dl_vlan_pcp=0x00,label_src=55,label_dst=65,nw_src=10.0.0.3,nw_dst=10.0.0.2,nw_tos=0x00,icmp_type=0,icmp_code=0,actions=output:3
cookie=0, duration_sec=1s, duration_nsec=162000000s,flow_id=523,table_id=0, priority=65535, n_packets=1, n_bytes=98,idle_timeout=10,hard_timeout=30,icmp,in_port=5,dl_vlan=0xffff,dl_vlan_pcp=0x00,label_src=55,label_dst=73,nw_src=10.0.0.3,nw_dst=10.0.0.10,nw_tos=0x00,icmp_type=0,icmp_code=0,actions=output:2

Tabla 6. Estructura de la *flow table* del nodo switch R3 para paquetes ICMP

Los nodos conmutan las etiquetas en espacios de tiempo programados. Capturamos tráfico en el enlace del MN con wireshark, y realizamos un análisis pormenorizado del mismo con tcptrace herramienta que proporciona gran cantidad de información, información de rendimiento, tiempos transcurridos, bytes/segmentos tanto enviados como recibidos, análisis de RTT, retransmisiones, tamaños de ventana, información de throughput, packet loss. Con toda la información, se elaboran las gráficas mediante jplot.

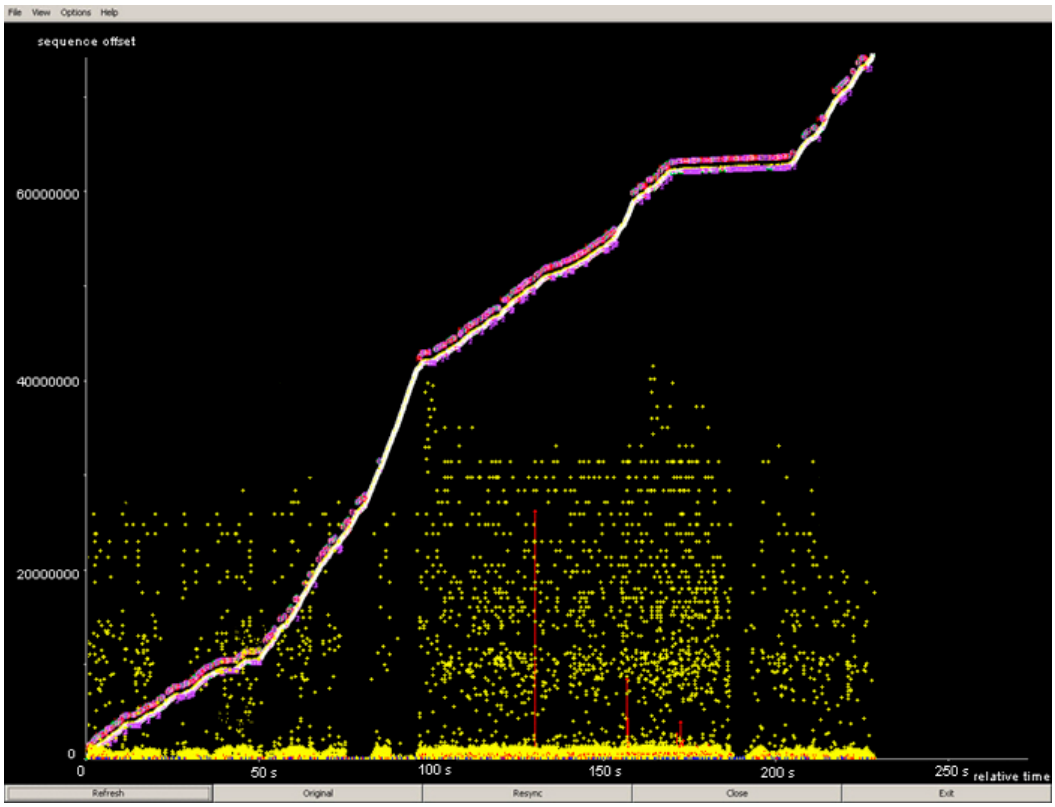

Figura 24. Números de secuencia y throughput en función del tiempo

La figura 24 ilustra el número de secuencia y el throughput en función del tiempo. Se puede ver como la pendiente de la curva del número de secuencia es diferente para cada uno de los dominios administrados. Esto se debe a la relación inversa del rendimiento de TCP con el retardo de ida y vuelta de la ruta que sigue el paquete. La pendiente representa la perdida de paquetes durante el handover del MN. El porcentaje es inferior al 1,45%. Para ello, representamos los números de secuencia que recibimos, y el throughput en función del tiempo.

El gráfico muestra seis valles en los tiempos de 49, 100, 165, 210 sg. Los valles corresponden a las pérdidas de paquetes debido a la adaptación del tráfico cuando el MN realiza un handover, esto es el *efecto shaping*. Una vez realizados los handover los números de secuencia vuelven a aumentar. Fundamentalmente debido a que se consideran los enlaces sin pérdidas en la configuración de la simulación. Un efecto similar ocurre con el throughput. En los handovers cuando la pendiente de la recta del número de secuencia se reduce, el throughput se reduce debido a que los paquetes se pierden y deben ser retransmitidos de nuevo.

Solamente, pretendemos caracterizar la solución, no validar y estandarizar la solución, debido a que el mecanismo para generar tráfico y detectar la pérdida de paquetes no es ideal. Esto es debido a que la simulación se lleva a cabo en una única máquina virtual y la virtualización de nodos no es eficiente.

La figura 25 representa la latencia que corresponde la los handovers, aunque los valores son parecidos en todos los casos. El emisor en todos los casos detecta rápidamente la pérdida de paquetes y se recupera de ella sin apenas dar lugar a discontinuidades en la curva, para volver rápidamente a transmitir con normalidad.

Este análisis demuestra experimentalmente como el empleo de MPLS dentro de esquemas SDN mejora claramente el rendimiento de las arquitecturas de movilidad en términos de latencia. Por lo tanto, su uso es beneficioso para entornos donde el retardo es muy alto. Aunque debido al método de medida empleado, no es posible ofrecer resultados más precisos.

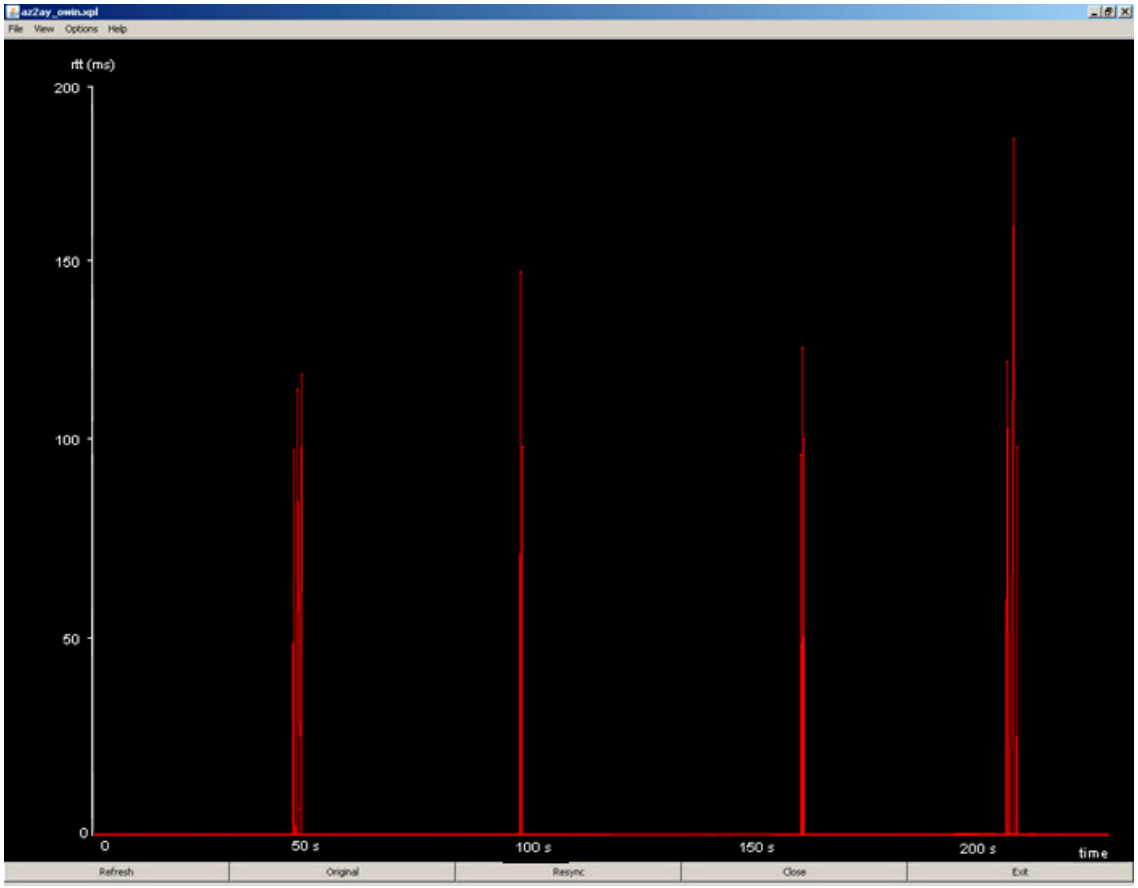

Figura 25. Latencia en el handover en un dominio IP/MPLS con OpenFlow

4.3. Evolución del protocolo OpenFlow en esquemas DMM

Las arquitecturas SDN soportan la gestión de la movilidad de forma transparente además de optimización de la ruta, permite también que un MN pueda enviar paquetes directamente al CN. En este apartado se describe una posible evolución del EPC utilizando esquemas SDN.

Actualmente los protocolos ofrecen escasos mecanismos para monitorizar el tráfico, también ofrecen escaso margen para aumentar el tiempo utilizado en señalizar la ruta de un paquete, o recuperarse ante fallos. Deben iterar sucesivas veces para conseguir rutas óptimas, por ejemplo en el caso de que exista congestión en un determinado enlace, deben señalizar enlaces alternativos. De ese modo, los estándares no proporcionan garantías de servicio para servicios en tiempo real debido a la ineficiente señalización. Por el contrario, la señalización si debe proporcionar garantías de servicio a dichos servicios. Servicios que por lo general adoptan conductas deterministas, acciones de entrada producen mismas acciones de salida. De ese modo se puede aplicar el principio número uno de la movilidad "prepararse antes de moverse".

Por otro lado, OpenFlow señaliza con mayor granularidad los flujos que atraviesan los nodos, conmuta paquetes de modo más rápido, y crea túneles en el plano de usuario con un menor coste desde el punto de vista de la señalización. Al mismo tiempo la red puede seguir operando de modo tradicional, a través de protocolos legados y utilizando los actuales dominios administrados.

El número de protocolos que soportan OpenFlow es reducido, si se compara con esquemas tradicionales. Actualmente están soportados los típicos protocolos de encapsulación de capa 3, como IPv4, IPv6, UDP, GRE, y GTP. Todos ellos realizan operaciones similares en cuanto a encapsulación y desencapsulación, procedimientos que pueden adaptarse a esquemas SDN, aunque algunos mejora la semántica del comportamiento del flujo.

La especificación OpenFlow 1.3 introduce nuevos campos de cabecera: TEID para GTP-U y KEY para GRE, esto es la capacidad para crear túneles genéricos en capa 3 para mapear los flujos a claves para diferenciar tuneles E2E respecto a otros. El mecanismo permite al controlador especificar un identificador de túnel que representa un GRE KEY [RFC2784][RFC2473]. La señalización la realiza el controlador a través de mensajes de tipo BPU donde se intercambian las KEY requeridas para identificar el GRE a través de extensiones del mensaje BPU. De ese modo, es posible señalizar el túnel PMIPv6 en lugar de utilizar el LMA y el MAG en el que negocian las claves KEY [49], las claves son suministradas por el controlador. Una solución similar puede aplicarse con túneles GTP en el que se crea un mismo túnel para una misma conexión, aunque separando el uplink, del downlink mediante valores TEID que son diferentes.

Utilizando OpenFlow 1.3 es posible implementar esquemas DMM de similar modo. En las soluciones basadas en host utilizando DSMIPv6, la señalización puede intercambiarse entre el MN y el controlador SDN. La idea que proponemos es que sea el controlador quien señalice los túneles genéricos entre los diferentes routers de acceso. El controlador pueda seleccionar o permitir la entrada del MN en los routers de acceso, a partir de la capacidad del router de acceso o preferencias del servicio del MN. La señalización se realiza a través de los mismos procedimientos del protocolo MIPv6, intercambiando mensajes de de tipo BU, y BA. Cuando un MN visita una red, este registra a través de eventos información en el controlador SDN. La solución basada en host causa un excesivo overhead en el MN. El controlador señaliza un túnel GRE Ipv6-to-IPv6, entre el MN y el anterior router de acceso, para ello crea un túnel bidireccional entre el anterior MAR y MN en la posición actual. Las nuevas sesiones se realizan a través del nuevo prefijo IPv6 asignado por el router de acceso. La ventaja del enfoque es que las sesiones no requieren de encapsulación ni señalización hasta que el MN no cambie de MAR, aunque por el contrario el MN esta involucrado en el proceso.

El diseño de una solución basada en red con PMIP, permite eliminar las entidades LMA y MAG. Entidades que se sustituyen por túneles genéricos instanciados por el controlador en los routers de acceso. La señalización la negocia el controlador con el router de acceso a través de mensajes de señalización típicos de PMIPv6, mensajes PBA, PBU. El controlador asigna los prefijos al MN dentro de los mensajes PBA y mantiene la sesión activa. Los rotuers de acceso informan al controlador cuando detectan el enlace de un nuevo MN, ya sea por tiggers de capa 2, o anuncios de ruta en capa 3. El controlador sabe si el MN procede de un handover, si procede de un handover señaliza el túnel PMIP entre los dos routers de acceso. El MN siempre se ve como un prefijo de red IPv6 único.

El desarrollo de la señalización con OpenFlow 1.3 obliga a introducir nuevas acciones push/pull, como pushIPv6, popUDP, pushGRE, etc. Cada de ellas puede desencadenar varias acciones push, esto es una diferencia respecto versiones anteriores de OpenFlow. Las acciones son similares a las soportadas en tecnologías VLANs y MPLS, permiten aislar flujos de tráfico de la infraestructura. Entre las ventajas de utilizar esquemas SDN en el EPC están, mejores convergencias en caso de fallo, la convergencia es inmediata. Mejores posibilidades de debugear o saber la trazabilidad del MN para personalizar el servicio. Posibilidades reales y eficientes de balanceo de carga.

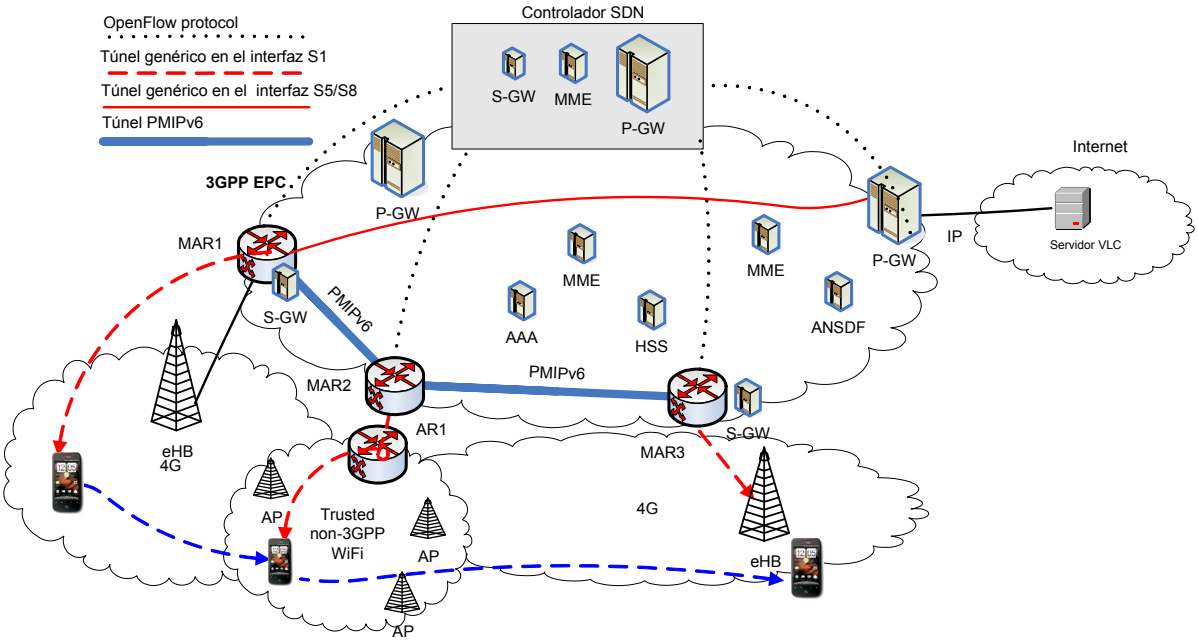

Figura 26. Arquitectura EPS utilizando esquemas SDN

4.4. Optimización de flujos de tráfico con OpenFlow

En este apartado realizamos un estudio teórico de las técnicas de optimización de flujos a través de esquemas SDN OpenFlow. Los flujos de tráfico caracterizados por paquetes pequeños presentan gran overhead en la infraestructura, servicios por lo general muy sensibles al retardo, y con requisitos de tasa de bit. Servicios como juegos online, o VoIP que para conseguir interactividad es necesario el envío de paquetes pequeños. Por lo general, el tamaño medio de un paquete para un servicio de juegos on-line es en media de 80 bytes, y el tamaño de un paquete VoIP es el múltiplo del tamaño de la muestra del códec, por ejemplo con un tamaño de muestra de 10 bytes, el payload puede contener 10, 20, 30 bytes circunstancia que genera gran overhead de red. El fenómeno es especialmente sensible cuando se comparte tráfico con otros tipos de tráfico. Hay que tener en consideración el tamaño de las cabeceras. En IPv4 el tamaño es de 40 bytes: 20 IPv4 + 8 UDP + 12 RTP, y en IPv6 60 bytes: 40 IPv6 + 8 UDP + 12 RTP. El fenómeno empeora con los 20 bytes de la cabecera TCP. Por otro lado, existen campos

de cabecera que aparecen en todos los paquetes y no cambian durante la ruta virtual que sigue un paquete dentro de un dominio administrado SDN.

En ese sentido se desarrollan mecanismos de compresión de cabecera como los algoritmos (RFC 3095), o Van Jacobson (RFC1144) que demuestran que 5 octetos de cabecera son suficientes para transmitir toda la información. Existen situaciones de tráfico interactivo y transferencias masivas de datos, donde es posible llegar a niveles de compresión incluso mayores de 3 octetos. Los campos que varían a lo largo de la conexión se sustituyen por incrementos respecto al valor del mismo campo del segmento inmediatamente anterior, de ese modo por ejemplo, el número de secuencia de un segmento recibido se calcula a partir del segmento anterior sumándole el valor incremental que proporciona el nuevo segmento. Se consigue que los algoritmos de compresión situados en el compresor únicamente envíen los campos de cabecera necesarios para que el descompresor pueda reconstruir la cabecera original a partir del paquete recibido y el contexto aprendido por el nodo. De ese modo se aumenta el throughput de la transmisión. Este hecho mejora las expectativas del usuario sobre el rendimiento de las comunicaciones. Aunque, un error en uno de los paquetes se propaga y afecta a todos los paquetes posteriores a él. Por esta razón, el algoritmo de compresión debe estar preparado para detectar posibles situaciones de error y recuperarse ante ellas.

Los mecanismos de compresión optimizan el overhead, aunque no consiguen mantener sincronizado el contexto, información para reconstruir las cabeceras en cada nodo. Si se pierde algún paquete que lleva información completa de cabecera, en el origen y en el destino no coinciden los valores de contexto valores como número de secuencia, o las marcas de tiempo en RTP, y tiene que pasar algún tiempo para recuperarlo debido a que se recupera cuando se envía un paquete sin comprimir. Por ejemplo, ante un error de contexto, el desencapsulador descarta la trama y envía al descompresor una señal para indicarle que debe desechar todos los paquetes hasta que reciba uno sin comprimir es decir, con valores absolutos y no relativos, del número de secuencia, o número de reconocimiento. El receptor se queda esperando a que envíen el paquete no comprimido.

Puesto que el descompresor en recepción no envía ningún tipo de señal al compresor en emisión, todo queda en manos del TCP emisor que al detectar la expiración del temporizador de retransmisión del segmento erróneo retransmite el segmento. El compresor en emisión detecta que el número de secuencia del nuevo paquete es menor o igual que el del último número de secuencia enviado, de ese modo esta obligado a transmitir el paquete sin comprimir para sincronizarse con el receptor.

Para solucionar, los problemas de contexto y de latencia se desarrolla *Enhanced Compressed RTP* (ECRTP), el procedimiento mejora la desincronización del contexto, modificando el modo en que cada nodo sincroniza el contexto. Para ello, los campos se actualizan periódicamente y alternan entre ellos, tanto los campos totalmente comprimidos, como los que no se comprimen o se comprimen de manera diferencial. Los campos de cabeceras se dividen en estáticos, aleatorios, y delta. Los estáticos se envían sólo en las primeras cabeceras. Los aleatorios no se comprimen. Los campos delta se van incrementando de uno en uno y se codifican con menos bytes que el campo original. (RFC 4170). La cabecera comprimida incluye como mínimo dos bytes a modo de máscara que indica los campos de la cabecera que se transmiten a continuación, y el checksum TCP de la cabecera original necesario para la detección de errores en la descompresión.

El IETF para la multiplexación de flujos RTP, define el estándar *Tunneling Multiplexed Compressed RTP* (TCRTP), que más bien es una recomendación para utilizar otros estándares que pueden comprimir diferentes cabeceras, por ejemplo comprimir solo IP+UDP, TCP, o ninguna. En primer lugar se usa ECRTP para comprimir las cabeceras IP, UDP y TCP. Posteriormente se usa el protocolo PPPMux (PPPMultiplexing) utilizado para multiplexar varios paquetes pequeños en un único paquete. El paquete comprimido se envía mediante

Point-to-point Protocol (PPP), un túnel *Layer 2 Tunneling Protocol* (L2TP), ó IP. El uso de tunelado permite utilizar ECRTP extremo a extremo, evitando así la necesidad de aplicarlo en cada router de la ruta, y utilizar técnicas de multiplexación. Las técnicas de compresión son independientes de la multiplexación.

Si se analizan los campos de las cabeceras de los principales protocolos TCP, IP, UDP, RTP, se observa que para una transmisión concreta, gran parte de los campos de las cabeceras se mantienen constantes, son deducibles o no varían. La situación se ve favorecida en dominios SDN para flujos donde los routers de entrada y salida reconstruyen cabeceras a partir de información contenida en la *flow table* que pertenecen a la estructura de datos del protocolo OpenFlow y accesibles a través del campo FID (*OFPXMT_OFB_IPV6_FLABEL*). El mecanismo propuesto aumenta la eficiencia de overhead, y añade solamente un pequeño retardo, que corresponde al tiempo de encolado en el multiplexor, y la preparación de cabeceras en el nodo de entrada y salida.

IPv6 (60 bytes)	PPPMux (2 bytes)	FID (3bytes)	IPv6 (3 bytes)	TCP (3 bytes)	Datos de usuario	Checksum (4 bytes)	PPPMux (2 bytes)	FID (3bytes)	IPv6 (3 bytes)	TCP (3 bytes)	Datos de usuario	Checksum (4 bytes)

Figura 27. Multiplexado y compresión de cabeceras optimizadas para TCRTP OpenFlow

La posibilidad de encapsular trafico multiplexado comprimido con compresión de cabecera tiene un impacto directo sobre los servicios portadores E2E. (Ver figura 28). Conseguimos un ahorro importante en la tasa de bit, de ese modo evitamos comprimir los campos delta, y también evitamos los problemas de sincronización en cada nodo. El inconveniente de la solución es el coste en el procesado, y adaptación del flujo de tráfico.

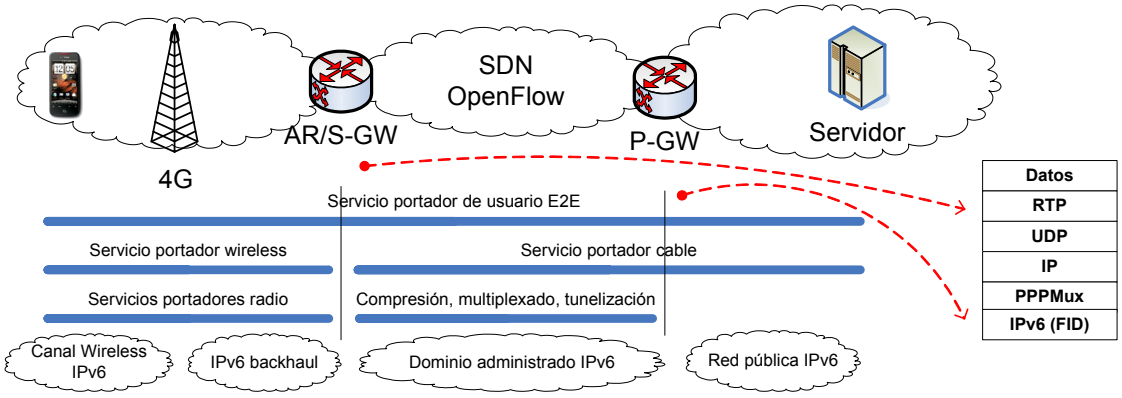

Figura 28. Arquitectura de un servicio portador de usuario E2E

Para poner en contexto las afirmaciones, se realiza una simulación con el codec G.729a con dos muestras por paquete. Esto supone un tiempo de paquetización de 25 ms, que incluye 10 ms por cada muestra, más otros 5 ms de look-ahead, esto es muestras del futuro para estimar mejor la señal de audio y poder así lograr una compresión mayor en el codificador. A menor tamaño de paquete menor retardo permitido auque se debe llegar a un compromiso. Cada flujo envía un paquete cada 20 ms, siendo el paquete de 20 bytes de payload más 40 bytes que corresponden a las cabeceras IP, UDP y RTP. Por tanto, el ancho de banda a nivel IP es de 24 kbps (60 bytes / 20 ms). A nivel Ethernet el ancho de banda será de 29,6 kbps, pues se añaden otros 64 bytes a cada paquete. Los juegos on-line generan 60 paquetes por segundo, con un tamaño de 80 bytes. Por lo que el ancho de banda es de unos 40,7 kbps por jugador.

Para obtener el ahorro de ancho de banda calculamos la diferencia entre el ancho de banda obtenido por flujos TCRTP, con flujos RTP nativos. La figura 29 representa ese ahorro. El multiplexado TCRTP siempre lleva asociado ahorro de ancho de banda respecto a RTP nativo hasta el 60% de ahorro para 20 paquetes multiplexados. El ahorro es significativo a partir del sexto flujo multiplexados, aunque posteriormente no aumenta debido al tamaño de paquete final. La compresión de cabeceras consigue aumentar el ahorro de ancho de banda. La razón es que el algoritmo de compresión actúa sobre los 60 bytes de cabecera de cada paquete IPv6,

aunque luego se añadan 15 bytes de cabecera común, los efectos se compensan, porque el algoritmo de compresión de cabecera IPv6 reduce en más de 50 bytes la cabecera original.

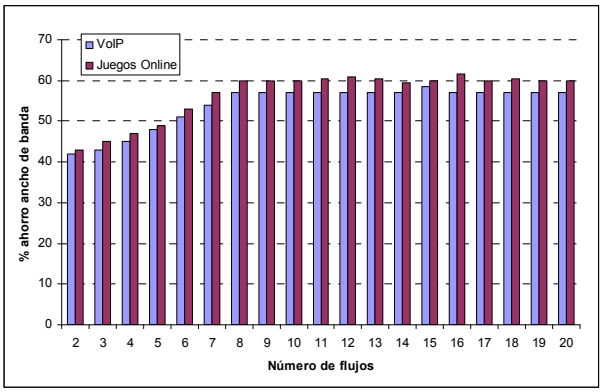

Figura 29. Número de paquetes multiplexados con compresión y ahorro de cabecera

La figura 30 representa el ahorro de ancho de banda en flujos multiplexados sin compresión. El mecanismo proporciona un buen rendimiento, aproximadamente igual a los mecanismos de compresión. De lo que se deduce que es mejor multiplexar paquetes que comprimir. La multiplexación reduce el número de paquetes por segundo, en una magnitud del orden del número de flujos multiplexados. Los buffers amortiguan el número de paquetes por segundo que generan, pero a mayor capacidad de buffer mayor retardo, esto es debido a que el buffer penaliza paquetes grandes, los descarta en caso de que la probabilidad de llenado del buffer sea alta, en ese caso hay que retransmitirlos. También hay que destacar a eficiencia del router para procesar paquetes, debido a que influye significativamente en el retardo, menos paquetes a procesar menor cuello de botella, paquetes pequeños generan gran cantidad de paquetes en el buffer. Para 20 flujos RTP nativos sin multiplexar el tamaño medio es de 89 bytes, mientras que al usar TCRTP se obtienen tamaños de paquete medios de 686 bytes.

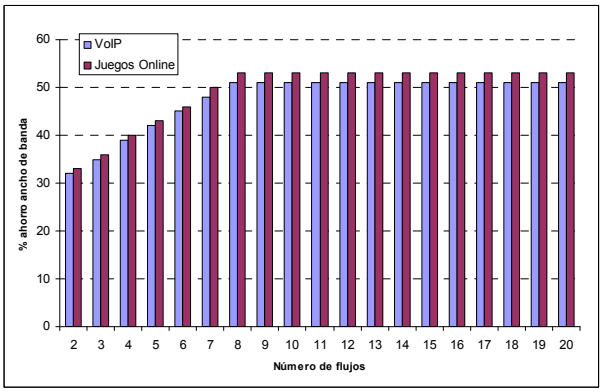

Figura 30. Número de flujos multiplexados sin comprimir vs ancho de banda

La figura 31 representa la agrupación de flujos RTP en un túnel TCRTP con diferente tamaño de paquete. Las asíntotas determinan el ancho de banda, aunque al tener asíntotas la mejora es escasa cuando el número de flujos multiplexados simultáneos aumenta. El ancho de banda se reduce muy poco cuando aumentamos el número de flujos multiplexados. Por tanto, si tenemos un gran número de flujos a multiplexar, la mejor solución no es agruparlos en un solo túnel puesto que el ahorro de ancho de banda será escaso, en cambio el tamaño de los paquetes puede crecer excesivamente. La multiplexación requiere de un compromiso entre ancho de banda, tamaño de paquete. Tal como se observa en la imagen la multiplexación no es óptima con cierto tamaño de paquete. Los paquetes pequeños sin overhead se adaptan mejor al ancho de banda del enlace, mientras que los paquetes grandes se descartan en un alto porcentaje, ya que tienen una mayor probabilidad de no caber en el buffer del router. El límite del paquete lo define la MTU, de ese modo, dependiendo de la red, es mejor utilizar paquetes pequeños que

grandes, siempre depende de la tasa de bit del enlace. Por esa razón, tamaños de paquetes grandes ocasionan mayores retardos.

Esto es debido al RTT, esto significa que si los paquetes son de gran tamaño, los valores adoptados por esta variable son elevados y, en consecuencia la expiración del temporizador de retransmisión tras una pérdida no detectable mediante reconocimientos tarda más en producirse. Por otro lado, cuanto mayor es la MTU del enlace, más tarda la ventana del emisor en ensancharse, debido a que los reconocimientos tardan más en llegar.

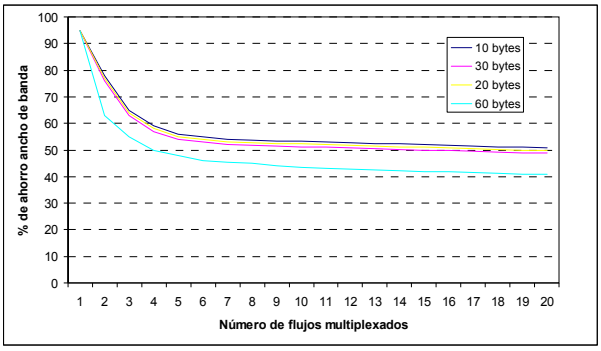

Figura 31. Ancho de banda respecto al tamaño de paquete multiplexado

Por lo tanto, la conclusión es que para ahorrar ancho de banda, la mejor opción consiste en multiplexar los flujos entre los routers de entrada y routers de salida en varios flujos con tamaños de paquete menores que la MTU del enlace. Buscando siempre un compromiso entre el tamaño de paquete, debido a que influye en la probabilidad de descarte en el buffer del router, a mayor tamaño mayor probabilidad. En segundo lugar, hay que tener en cuenta el tamaño del buffer, ya que esta directamente relacionado con la latencia del enlace, y la probabilidad de descarte. Finalmente, paquetes pequeños no optimizan los recursos del enlace, paquetes grandes causan excesivo retardo, debido al descarte y la fragmentación, con lo que la pérdida de un fragmento implicará la retransmisión del paquete entero.

5. Conclusiones y líneas futuras

El trabajo analiza el protocolo PMIPv6 y sus extensiones propuesto por el IETF como mecanismo para mantener la continuidad en la sesión del MN, continua con su evolución debido a la necesidad de adaptar el tráfico a nuevas arquitecturas que aparecen alrededor de LTE, analiza como es adoptado por el 3GPP para proveer soluciones de movilidad debido a las ineficientes soluciones desarrolladas por el organismo para mantener la continuidad en los servicios de banda ancha ofrecidos al usuario con requisitos de QoE.

La propuesta de trabajo propone un marco real para escalar servicios, consigue abstraer recursos de red para modelar servicios en términos KPI's, uno de los principales beneficios del enfoque es la simplificación de la red móvil. Son modelos que solucionan los problemas actuales dentro de arquitecturas monolíticas, arquitecturas difíciles de escalar, securizar y gestionar. El trabajo profundiza como es posible desplegar servicios eficientes haciendo uso de la virtualización utilizando para ello esquemas SDN capaces de señalizar la ruta por la que conmuta el flujo. Los esquemas SDN son mecanismos excelentes para escalar servicios, aunque no son la panacea a las ineficiencias de señalización actuales. No virtualizan la red, esto es separan la arquitectura lógica de la física, llevan a cabo una separación de la señalización respecto los datos. Los esquemas SDN tampoco son capaces de optimizar la infraestructura para paquetes pequeños. Para ello, se modelan los paquetes pequeños mediante flujos contenedores, multiplexando de paquetes con compresión de cabecera. Entre los beneficios de la multiplexación cabe destacar el ahorro de ancho de banda y la reducción del número de paquetes por segundo siempre para servicios caracterizados por su pequeño tamaño de paquete.

El paradigma de la conmutación aporta mayor escalabilidad, y mayor control de la QoS en redes. Desde el punto de vista del operador aporta un mayor control sobre la TE, el accounting, la gestión de recursos y la movilidad desde un enfoque basado en red. Siendo MPLS, el ejemplo que engloba todas estas características, y que ha sido analizado ampliamente en el presente trabajo. Los operadores prefieren el servicio prestado, en lugar de la tecnología necesaria para prestar el servicio, bajo estas circunstancias, conseguir aumentar QoE, tiene un gran impacto en la lealtad de los clientes.

Para modelar la infraestructura se tuvo en cuenta los siguientes factores, el primer factor de diseño se refiere al propósito del sistema móvil, maximizar el número de usuarios capaces de cursar el mayor volumen de tráfico posible. En segundo factor el medio de comunicación utilizado por el sistema móvil, el medio radioeléctrico es en sí mismo es un medio compartido, por lo que es preciso establecer unas adecuadas reglas para que los usuarios puedan comunicarse sin colisionar entre ellos, reglas que llamamos QoS. Estas reglas determinan el éxito en el despliegue de servicios de banda ancha.

Por otro lado, los recursos asignables por el medio son limitados, por ejemplo los dispositivos móviles tienen una importante limitación de la batería, el interfaz de red consume el 15% de la batería, el envío y la recepción de paquetes consume batería. Por ese motivo, el tráfico de señalización debe ser el mínimo posible, también debe aislar al MN de los mecanismos de señalización. Para ello, se propone señalizar el tráfico mediante tuplas únicas que identifican el flujo en cada nodo, donde la red lleva a cabo su gestión.

Actualmente la realidad es diferente, la promesa de tráfico por parte de los operadores obliga a buscar nuevos mecanismos para estructurar el tráfico. De ese modo, el despliegue de servicios de banda ancha exige flexibilizar el diseño de las arquitecturas bajo dos puntos de vista. En primer lugar es necesario flexibilizar la cobertura y en segundo lugar tener en cuenta como se utiliza el espectro radioeléctrico. En este escenario, surgen las small-cells con el objetivo de mejorar la experiencia de usuario y la capacidad del sistema en términos de tasa de bit. El enfoque introduce problemas de movilidad en el MN, handovers y la gestión de la QoS. Para ello, el trabajo analizó diferentes enfoques basados en arquitecturas DMM, que junto a las posibilidades multi-interfaz del MN y la inteligencia de la infraestructura consiguen disminuir la latencia. Soluciones que estructuran de modo inteligente el tráfico dentro de la macrocelda y consiguen escalar el servicio optimizando los recursos de la infraestructura.

Las futuras investigaciones deben focalizarse en la preparación de handovers predictivos en entornos cross-layer, para optimizar los flujos de tráfico a través de las redes de acceso dentro de sistemas heterogéneos multi-interfaz. Las arquitecturas de nueva generación se caracterizan por ser ecosistemas donde se solapan coberturas entre diferentes tecnologías. De ese modo, en el futuro debe ser posible construir soluciones en que los servicios interactúen directamente con los procedimientos de handover a través de esquemas SDN, por ejemplo para un flujo de datos en tiempo real, se deberá poder realizar un handover predictivo basándose en la configuración de la fuente el stream. Esto significa por ejemplo la capacidad de adaptar los parámetros del codec a su nuevo entorno, o proporcionar la interfaz y la tecnología por donde se va a proporcionar el nuevo flujo. También la posibilidad en el caso de que exista congestión de red encontrar en modo de cambiar los parámetros del codec del flujo. Para ello se deberá expandir el paradigma *match/action* en entornos SDN y conseguir *actions* más complejas que junto con los módulos hardware del nodo switch permitan encriptar, transcodificar, o inspeccionar flujos en tiempo real.

Los usuarios están reconociendo y entendiendo todas las promesas ofrecidas por la banda ancha. Los operadores por el contrario están buscando nuevas fuentes de tráfico. Tráfico encapsulado dentro de un servicio. De ese modo, un servicio se puede definir como una transacción que quizás necesite administrarse, almacenarse o incluso ser virtualizada. La transacción tiene que ser capaz de transformar las necesidades de usuario en un conjunto de

métricas QoS específicas de red, a través de un determinado SLA. Resaltar por su importancia los parámetros de seguridad, ya que también son parte de los tributos de una transacción.

Por lo tanto, consideramos la virtualización como un mecanismo excelente para proveer servicios cada vez más exigentes en términos de throughput y latencia por parte del usuario. El mecanismo permite crear recursos virtuales desde recursos físicos de red. La virtualización permite desacoplar las diferentes capas de la arquitectura e infraestructura de modo que es posible un aislamiento de los diferentes flujos de datos. Se consigue un aumento del rendimiento, obteniendo mayor escalabilidad y flexibilidad. La solución mejora el despliegue de nuevas arquitecturas de redes, y optimiza los protocolos de comunicación, permitirá en un futuro crear instancias de servicios en el *cloud computing*.

Para concluir, actualmente existen métricas, y mecanismos que proporcionan una QoE a partir de métricas QoS, aunque existe cierta dificultad para trasladar las métricas objetivas a servicios susceptibles de ser evaluados por el usuario. En el futuro las investigaciones deben centrarse en cuantificar y extender las clases de servicio en términos QoS dentro de entornos convergentes de redes integradas que permitan definir las reglas con las que aprovisionar dinámicamente los recursos de red.

Bibliografía

[1] 3rd Generation Partnership Project (www.3gpp.org).

[2] http://en.wikipedia.org/wiki/Tragedy_of_the_commons (Hardin G. 1968).

[3] G. Giaretta, "Interaction between PMIPv6 and MIPv6: Scenarios and Related Issues," draft-ietf-netlmm-mip-interactions-06, May 2010.

[4] Mobility Support in IPv6. IETF RFC 6275.

[5] UMTS Forum. (June, 2010). Recognising the Promise of Mobile Broadband. White Paper.

[6] http://www.ericsson.com/res/docs/whitepapers/differentiated_mobile_broadband.pdf.

[7] "General Packet Radio System (GPRS) Tunneling Protocol User Plane (GTPv1-U)", 3GPP TS 29.281, Release 9, 2010.

[8] "Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control Plane (GTPv2-C)", 3GPP TS 29.274, 2010.

[9] 3GPP TS 29.275: "Proxy Mobile IPv6 (PMIPv6) based Mobility and Tunnelling protocols; Stage 3".

[10] T. Salam, S. Mushtaq, T. Khalid, M. Ali, and M. Amin, "Efficient vertical handover approaches for increased user satisfaction in next generation networks," presented at the 2011 International Conference on Computer Networks and Information Technology (ICCNIT), 2011, pp. 233–239.

[11] Dutta, A. et al., "Media-independent pre-authentication supporting secure interdomain handover optimization" IEEE Wireless Communications, vol.15, no.2, pp.55-64, April2008.

[12] Arkko, J., Vogt, C., & Haddad, W. (2007. May). Enhanced Route Optimization for Mobile IPv6. IETF RFC 4866.

[13] IETF RFC 4140, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)", August 2005.

[14] Aguiar R., Banchs A., Bernardos C. J., Calderón M., Liebsch M., Melia T., Pacyna P., Sargento S. and Soto I." Scalable QoS-aware Mobility for Future Mobile Operators" (2006) .

[15] 3GPP TS 36.304 "User Equipment (UE) procedures in idle mode (Release 8)".

[16] A Performance Comparison of Mobile IPv6, Hierarchical Mobile IPv6, Fast Handovers for Mobile IPv6 and their Combination, 2003).

[17] 3GPP TS 23.401, General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access, September 2012.

[18] G. Tsirtsis et al., "Traffic Selectors for Flow Bindings," IETF RFC 6088, Jan. 2011.

[19] http://tools.ietf.org/html/draft-vonhugo-multimob-dmm-context-02.

[20] TS 24.312, Access Network Discovery and Selection Function (ANDSF) Management Object (MO): http://www.3gpp.org/ftp/Specs/html-info/24312.htm.

[21] Zhenhua, W., Qiong, S., Xiaohong, H., & Yan, M. (2010). IPv6 end-to-end QoS provision for heterogeneous networks using flow label. 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT), (old.: 130-137).

[22] Distributed Mobility Anchoring (http://tools.ietf.org/html/draft-seite-dmm-dma-01).

[23] Traffic Selectors for Flow Bindings (http://tools.ietf.org/html/rfc6088).

[24] Mobility Practices and DMM Gap Analysis. http://tools.ietf.org/html/draft-zuniga-dmm-gap-analysis-03.

[25] Requirements for Distributed Mobility Management (http://tools.ietf.org/html/draft-ietf-dmm-requirements-03).

[26] Chan, H., "Distributed Mobility Management with Mobile IP", Proceedings of IEEE International Communication Conference (ICC) Workshop on Telecommunications: from Research to Standards, June 2012.

[27] Local Prefix Lifetime Management for Proxy Mobile IPv6 (http://tools.ietf.org/html/draft-korhonen-dmm-local-prefix-00).

[28] Calyam P., Lee Chang-Gun, Characterizing voice and video traffic behavior over the Internet, In Advances in Computer Science and Engineering: Reports, Imperial College Press, Proceedings of ISCIS 05, Turkey 1 (2005).

[29] Multiple Care-of Addresses Registration http://tools.ietf.org/html/rfc5648.

[30] http://www.openflow.org/documents/openflow-spec-v1.1.0.pdf.

[31] MPLS-TP Pseudowire Configuration using OpenFlow 1.3 draft-medved-pwe3-of-config-01 (http://tools.ietf.org/html/draft-medved-pwe3-of-config-01).

[32] Logical Interface Support for multi-mode IP Hosts (http://tools.ietf.org/html/draft-ietf-netext-logical-interface-support).

[33] Controlador POX (http://www.noxrepo.org/pox/about-pox).

[34] Manual de referencia Mininet Python API (http://mininet.org/api/index.html).

[35] Click Modular Router Project (http://www.read.cs.ucla.edu/click/click).

[36] Especificación para las mejoras en accesos non-3GPP (3GPP TS 23.402 v12.0.0) .

[37] Requerimientos funcionales para estructuras "small-cells" (3GPP TR 36.932 v12.1.0).

[38] Mejoras en movilidad en redes heterogéneas (3GPP TS 36.839 v111.0).

[39] Policy and Charging Control Architecture (3GPP TS 23.203 v12.1.0).

[40] L. Stewart, P. Branch, Quake4, Map: q4dm1, 7players, 20Jul2006. Centre for Advanced Internet Architectures SONG Database, http://caia.swin.edu.au/sitcrc/song/files/quake4_310706_1_q4dm1_7_fragment.tar.gz.

[41] OpenFlow1.3 (https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.3.0.pdf).

[42] European Telecommunications Standards Institute (http://portal.etsi.org).

[43] ITU-T, ITU-T Recommendation G.114, One-Way Transmission Time, Standard 1397 G.114, ITU-T (International Telecommunication Union, Telecommunication 1398 Standardization Sector), February 1996.

[44] VoIP: A comprehensive survey on a promising technology Stylianos Karapantazis *, Fotini-Niovi Pavlidou.

[45] Online access of webpage, http://www.videolan.org/doc/vlc-userguide/en/ch01.html, on 12th March 2010.

[46] Giovanni Gualdi; Rita Cucchiara; Andrea Prati, "Low-Latency Live Video Streaming over Low-Capacity Networks", Eighth IEEE International Symposium on Multimedia, 2006. ISM'06, Publication Year: 2006, Page(s): 449 – 456.

[47] Acceso Online videos test (http://media.xiph.org/video/derf/).

[48] Soldani, D., Li, M., Cuny, R.: QoS and QoE Management in UMTS Cellular Systems. Wiley, New York (2006).

[49] draft-muhanna-netlmm-grekey-option-00.txt.

[50] "Reinforcement learning for joint radio resource management in LTE UMTS scenarios." of N. Vucevic, J. Pérez-Romero, O. Sallent, y R. Agustı.

[51] Emulab (http://www.emulab.net).

[52] http://www.lteforum2013.com.